GENETICALLY ENCODABLE CHROMATIC LUMINESCENCE FOR NEURAL ACTIVITY MAPPING

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Mitchell Alan Pender

May 2019

GENETICALLY ENCODABLE CHROMATIC LUMINESCENCE FOR NEURAL ACTIVITY MAPPING

Mitchell Alan Pender, Ph. D., MBA

Cornell University 2019

The brain is the most complex organ in the human body. Composed of intricate overlapping networks existing on multiple spatiotemporal and philosophical scales, a comprehensive and unified understanding of the brain does not exist. The genomics revolution has enabled new approaches in scientific innovation through novel tool development. These genetically encodable tools may offer an improved and translatable approach in understanding the brain's complexity. Here, I present methods for engineering viral vectors which genetically encode chromatic photoproteins to transduce neurons in the mammalian cortex such that each neuron may have a unique spectral signature. The photoproteins generate luminescence in response to Ca^{2+} oscillations, a primary messenger in neuronal signaling. Optically separated luminescent signals serve as a platform for feature generation where machine learning ensembles are able to correlate signal features and predict responses to external stimuli. With analysis suggesting increased color heterogeneity improves predictor accuracy, I engineered novel photoproteins with enhance kinetic and chromatic properties. With continued optimization of indicator, detector, and algorithmic design, this approach could serve as a translatable platform for a human brain-machine interface.

BIOGRAPHICAL SKETCH

Mitchell Pender was born in 1990 and raised a “Yooper” in Iron Mountain, Michigan; a region so isolated that the state legislature had to pass House Bill No. 4995 mandating that all future U.S. maps remember to include the Upper Peninsula. In 2009 he attended Albion College, studying organometallic synthesis and spectroscopy under Vanessa P. McCaffrey. He graduated in May 2013 at the top of his class, earning numerous regional and national awards in analytical chemistry. In August 2013 he began the pursuit of a doctoral degree in Biomedical Engineering at Cornell University. Discovering the power of genetic technologies through his research, he joined the efforts of a third-generation DNA nanopore sequencing start-up company, Esper Biosciences. In 2017, entrepreneurial spirit influenced him to pursue and complete his MBA at the SC Johnson College of Business. He spent the following summer as a strategic analyst for Corning Incorporated. Unable to escape his passion for scientific discovery, he returned to Cornell in August 2018 to complete his PhD. In June 2019 he will join Corning’s Emerging Innovations Group as a project manager conducting competitive and market intelligence research for novel technologies.

Dedicated to my family: Dan, Cathy, and Josh

ACKNOWLEDGMENTS

Leah Pagnozzi, Nathan Gibson, Yingchao Wang, Karen Lin, Heley Ong, Eva Ding, Xin Li, Amanda Bares, Jason Jones, Jeffrey Mulligan, John Foo, Yu-ting Cheng, Elizabeth Wayne, Sung-Ji Ahn, Jiahn Choi, David Small, Oliver Brocko, Lutz Finger, Risa Mish, Marty Curran, Anis Fadul, Richard Miller, Daniel Orr, Mark Soulliere, Jan Lammerding, Steven Adie, Nozomi Nishimura, Christopher B. Schaffer, Stephen Macknik, Mathew Paszek, Warren Zipfel, Becky Williams, Michael Kotlikoff, Katie Barajas, AG Colinco, Alex Van Slyke, Max Kushner, Jonathan Alden, Joshua Alden, George Calvey, Lois Pollack, Ben Looker, Liz Wills, Colin Parish, Joel Baines, Wendy Weichert. Government support: NSF 1706761, NSF 1453339, and NIH EY028391.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

BMI	Brain machine interface
MPM	Multiphoton microscopy
MEG	Magnetoencephalography
fMRI	Functional magnetic resonance imaging
PCA	Principal components analysis
AEQ	Aequorin
CTZ	Coelenterazine
FP	Fluorescent protein
GFP	Green fluorescent protein
RET	Resonance energy transfer
CRET	Chemiluminescent resonance energy transfer
BRET	Bioluminescent resonance energy transfer
CMOS	Complementary-metal-oxide semiconductors
FOV	Field of view
PTZ	Pentylenetetrazol
AAV	Adeno-associated virus
2PEF	Two-photon excited fluorescence
RMSE	Root mean squared error
PC	Principal components
DWT	Discrete wavelet transform

DFT	Discrete Fourier transform
FFT	Fast Fourier transform
PMT	Photomultiplier tubes
MWT	Mother wavelet
CWT	Continuous wavelet transform
NA	Numerical aperture
MAD	Median absolute deviation
RWE	Relative wavelet energy
BBB	Blood brain barrier
GDP	Gross domestic product
RLU	Relative luminescence output
AA	Amino acid
CP	Circular permutation
mNG	mNeonGreen
OBE	Obelin
AMAC	Aequorin macrodactyla

CHAPTER 1

1: Introduction and Organization of Dissertation

1.1 Introduction

The global economy is experiencing the start of a broad biotechnology revolution comparable to the industrial revolution of the 19th century, and to the Digital Age of today. This revolution is characterized by accelerated growth in biotechnology start-up companies, a shift from small to large molecule pharmaceuticals development, and the dynamic and growing presence of multi-omics bioanalytical techniques for advancing our understanding of human health. At the core of this revolution lies the development of technologies for understanding and manipulating gene function in living organisms, as well as advanced systems biological methods for detecting phenotypic changes and diagnosing disease.

As this era begins to demonstrate unprecedented advances for fields like genomic sequencing and metabolomics, tools to fully understand and recapitulate the complex nature of the human brain in an unrestricted environment remain elusive. Merge this intractable technological hurdle with notable growth in neurological disease and the call for a paradigm shift in neural activity recording has become ever more opportune. In this context, I have dedicated most of my graduate career towards building and testing optical and genetically encodable tools for probing in vivo physiology while demonstrating an entirely novel approach for understanding brain function. This dissertation attempts to describe my efforts in protein engineering, virology, imaging, signal processing, and machine learning towards building the framework for a human translatable brain-machine interface. Although the technology

is it its infancy, and human translation is a distant reality, the concepts described here are well poised to gain traction in the growing biotechnology revolution.

1.2 Organization of Dissertation

In 2013, the National Institutes of Health outlined a sizeable call for neuroscientists to develop new methods for understanding brain function. Chapter 2 presents an answer to their call with a novel combination of optical and genomic technologies that could serve as a future brain machine interface. An overview of fundamental design concepts are presented in **Figure 1.1**.

In Chapter 3, a framework for signal analysis utilizing the discrete wavelet transform is presented. Bioluminescent signals emerging from the rodent cortex (**Fig 1.1d**) are collected on a system optimized for scattered light detection (**Fig 1.1b**). Photomultiplier tubes detect spectrally multiplexed signals (**Fig 1.1a**), which are sampled for discrete wavelet deconstruction and feature generation.

Results from Chapter 3 motivate Chapter 4. Development and characterization of novel bioluminescent calcium sensitive fusion constructs are discussed to enhance function of the neural interface through improving kinetics and brightness while expanding the available color palette. Bioluminescent spectra for a range of the developed and existing indicators are introduced in **Figure 1.1g**.

In Chapter 5, development and characterization of novel circularly permutated (cp) variants of fluorescent protein mNeonGreen are presented. The cp-variants are inserted into binding domains of calcium or ATP and characterized in vitro and in vivo. Experiments are ongoing, but preliminary characterization is promising.

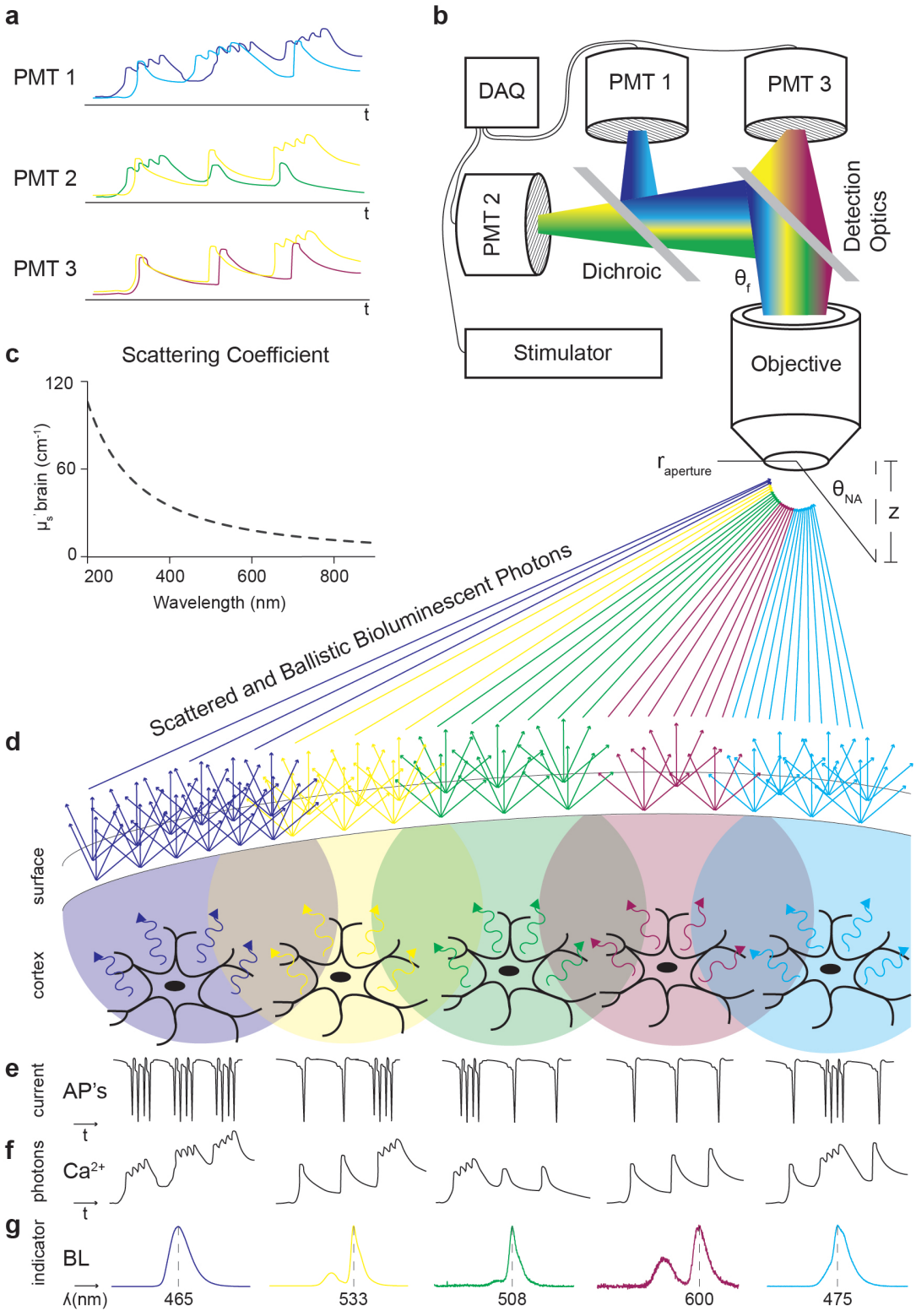


Figure 1.1 Overview of bioluminescence for neural activity mapping
a-g) Representative elements of the neural mapping technologies that are described throughout this dissertation. Chromatic indicators report neural activity via non-spatially resolved bioluminescence.

CHAPTER 2

2: Chromatic Bioluminescence as A Cellular Level Readout System of Neural Activity

2.1 Abstract

Next-generation brain machine interfaces (BMIs) for applications such as prosthetics require the ability to map activity of neural networks distributed over large areas, maintain the sensitivity and resolution to detect firing of a single neuron, and must be durable. Current BMIs based on implanted electrodes can record activity from individual neurons but cannot accurately record more than a few neurons per electrode and degrade over time. Our approach measures neural activity by light emission from calcium-sensitive bioluminescent proteins known as fluorescent protein-aequorins (FP-AEQs). We "barcode" neurons using a gene-therapy viral vector approach to express assortments of different color FP-AEQs. Each neuron may have stochastically varying amounts of each color, so that when a neuron fires it can emit light with a distinct spectral signature. We recorded bioluminescent light from an anesthetized mouse in response to a pentylenetetrazol induced seizure and whisker stimulation. Trajectories of intensity at different wavelengths showed different states before, during, and after a stimulus. Discrete wavelet decomposition and frequency domain analysis generate time relevant features which cluster to represent the existence of multiple possible activity states in response to stimuli. A framework and analysis is presented for utilizing bioluminescent signals to predict activity and suggests that color-coding increases the complexity of the information encoded relative to single-color reporters. This tool could

provide a long-term, robust BMI for patients with prosthetic devices, with limitations discussed.

2.2 Introduction

A brain-machine interface (BMI) provides a direct pathway for the brain to communicate with an external device, bypassing the need for peripheral motor or sensory involvement[1]. To do this, a BMI attempts to capture information encoded in spikes, electrical pulses that neurons use to communicate amongst each other. An indirect method, such as functional magnetic resonance imaging (fMRI), might measure changes in brain-blood flow that follow regional spiking events[2]. Direct methods utilize microelectrodes, microwires implanted near or within a cell body which are capable of recording the spiking amplitude and duration of individual neurons[2,3]. Through the control of a prosthetic, BMIs capable of high fidelity spike recording have the potential to restore function in a patient who has suffered from a paralyzing ailment such as a stroke or amputation[4]. Through advances in neural mapping technologies, BMIs can improve diagnosis and treatment of other neurological diseases through interrogating aberrant neuronal connectivity across multiple brain regions simultaneously and with genetic specificity[5,6]. To recapitulate these complexities and restore function with high fidelity, next-generation BMIs require cell resolved, spatially scalable, genetically targeted, portable, and minimally invasive recording techniques that must allow a patient to maintain the freedom of natural behavior[7].

Although there are many technologies under development, none of them meet the entirety of demands required for complex neural recording[7,8]. Implanted electrodes

are able to resolve neural activity with single cell resolution at ever increasing spatial scales and geometries, but still suffer from the triggering of tissue immune response, long-term recording instability, theoretical limitations on tolerable brain volume displacement, and inability to distinguish specific cell subtype[3,9]. Optical fluorescent techniques such as multiphoton microscopy (MPM) and fiber photometry, using genetically targetable indicators of calcium or voltage, continue to make strides in offering unprecedent access to cell resolved neural activity[10–13]. In the case of MPM, a juggling act between laser power, wavelength, scan speed, and indicator biophysics impose a trade-off between temporal and spatial resolution, typically worsening with imaging depth[14]. On one end of the spectrum optimized MPM instrumentation is far too bulky to be portable, while on the other end head-fixture devices for awake-imaging suffer from in-plane motion artifact and resulting signal corruption[15]. In the case of multi-channel fiber photometry, sampling bulk fluorescence in a small region provides superior temporal resolution but can lead to fluorescence excitation cross-talk between probe tips and sacrifices nearly all spatial information aside from the fibers known implant location[16].

Experimental neuroscientists often agree that the human brain varies amongst individuals, communicates via billions of neurons each with its own potentially unique spiking characteristics, and creates trillions of connections across large spatial scales. Scientists strive to support brain function hypotheses by relating specific behaviors to resolved function of neural circuits. There is often less consideration than necessary given to how the scope of neural recording scales with the complexity of behavior. For instance, the principal components analysis (PCA) of recordings from 99 locust

antennal lobe projection neurons in response to odor stimuli could have be explained principally in three-dimensional space[17,18]. This conundrum merits further efforts to record active cells from larger ensembles that provide the most information gain in explaining the system variance. At the growth rate of current technologies, we stand a century away from simultaneously observing single cells across large enough spatial scales to understand complex mammalian behavior[17]. While placing more electrodes onto increasingly biocompatible chips and sending more photons deeper into the cortex has advanced our understanding of the brain significantly, the rise of translational genetic technologies suggest that in order to build on that understanding it may be time to start exploring paradigm shifts in the recording of neural activity[19].

Conceptually, one can envision a genetically encodable probe that reports neural activity directly through biophysical changes in the cellular environment without the requirement of direct electrode contact or external laser excitation. This probe might emit radiation in response to changes in cellular physiology, such as bioluminescence in response to a change in cell calcium or membrane voltage[20,21]. In principle, we could record these responses from highly sensitive photodetectors with subdural positioning in a human patient[8]. Naumann et al. first demonstrated a non-imaging technique for recording neural activity in freely swimming zebrafish using a bioluminescent indicator, Aequorin (AEQ), from *Aequorea Victoria*[22]. Upon reconstitution with its luciferin substrate, coelenterazine (CTZ), Aequorin is capable of emitting light in response to the calcium transient that occurs subsequent to a neuron firing a series of action potentials[23,24]. In jellyfish, AEQ is naturally complexed with green fluorescent protein (GFP) which led to the creation of the fusion protein GFP-AEQ. This fusion protein

enables resonance energy transfer (RET) from AEQ to GFP during calcium binding, enhancing calcium dynamics while maintaining the ability to roughly localize transients through imaging GFP[25]. In zebrafish, Naumann et al. employed a large area photodetector to record GFP-AEQ bioluminescence from genetically specified neurons over long periods with single cell sensitivity; noting limitations on sensor and photon-detecting technologies[22]. Although further optimized fluorescent protein-aequorin (FP-AEQ) sensors have been engineered that range in wavelength and design, techniques that employ them in the mammalian brain for monitoring cell-resolved activity have yet to be developed[26–30].

While bioluminescent calcium sensors have excellent signal-to-noise ratio, their chemically produced light signal is rather weak and constrains its detection to photon counting modalities such as photomultiplier tubes (PMTs)[31]. Although these detection modalities enable emitted photons to be detected from a large field-of-view (FOV) with high temporal resolution, they generally sacrifice all spatial information aside from the physical location of the detector's collection head. Here, we introduce a novel technique that presents an opportunity to replace the spatial resolution lost to this non-imaging modality via genetically encoding neurons with unique spectral signatures. We describe an adeno-associated viral vector (AAV) gene therapy approach to spectrally label individual neurons such that each neuron may have stochastically varying amounts of spectrally diverse FP-AEQ reporters. In collaboration with Bares and Pender et al.[32–35] we utilized hyperspectral two-photon excited fluorescence (2PEF) microscopy to excite AAV mediated fluorescence in rodent cortex, validating our spectral mix in a small region. Using a similar 2PEF microscope's detection optics, we

block external laser excitation and sample spectrally separated calcium dependent bioluminescence from a live rodent cortex in two scenarios: pentylenetetrazol (PTZ) induced seizure activity of a rodent discretely expressing two colors of FP-AEQs and via whisker pad electrical stimulus driven activity of a rodent expressing a stochastic mix of three FP-AEQ colors. A combination of time-domain averaging and convolution revealed interesting temporal properties of the bioluminescent signals. Further analysis utilizing the discrete wavelet transform (DWT) and level dependent statistics generated large frequency band feature vectors which were then fed into a range of machine learning ensembles. Although experiments are ongoing, preliminary analysis suggests that color coding may increase the complexity of information relative to single color reporters and that stimulus driven luminescence in mouse cortex may code for multiple different activity states. Further exploration of spectrally multiplexed light-emitting neurons may offer solutions to hurdles facing next-generation BMIs and answers to the complexity required to understand neural function.

2.3 Results

2.3.1 AAV vectors produce stochastic patterns in mouse cortex *in vivo*

The simultaneous mixing and co-injection into rodent cortex of multiple AAVs, each coded to express a unique color of fluorescent protein was explored with hyperspectral 2PEF microscopy *in vivo* through a chronically implanted cranial window or ex-vivo in formaldehyde fixed histological slices. Using a custom-built hyperspectral 2PEF microscope with three different excitation sources (**Fig. 2.1a**), and four angle-tuned bandpass filters (**Fig. 2.1b**), 48-channel spectrally unmixed images of a mouse

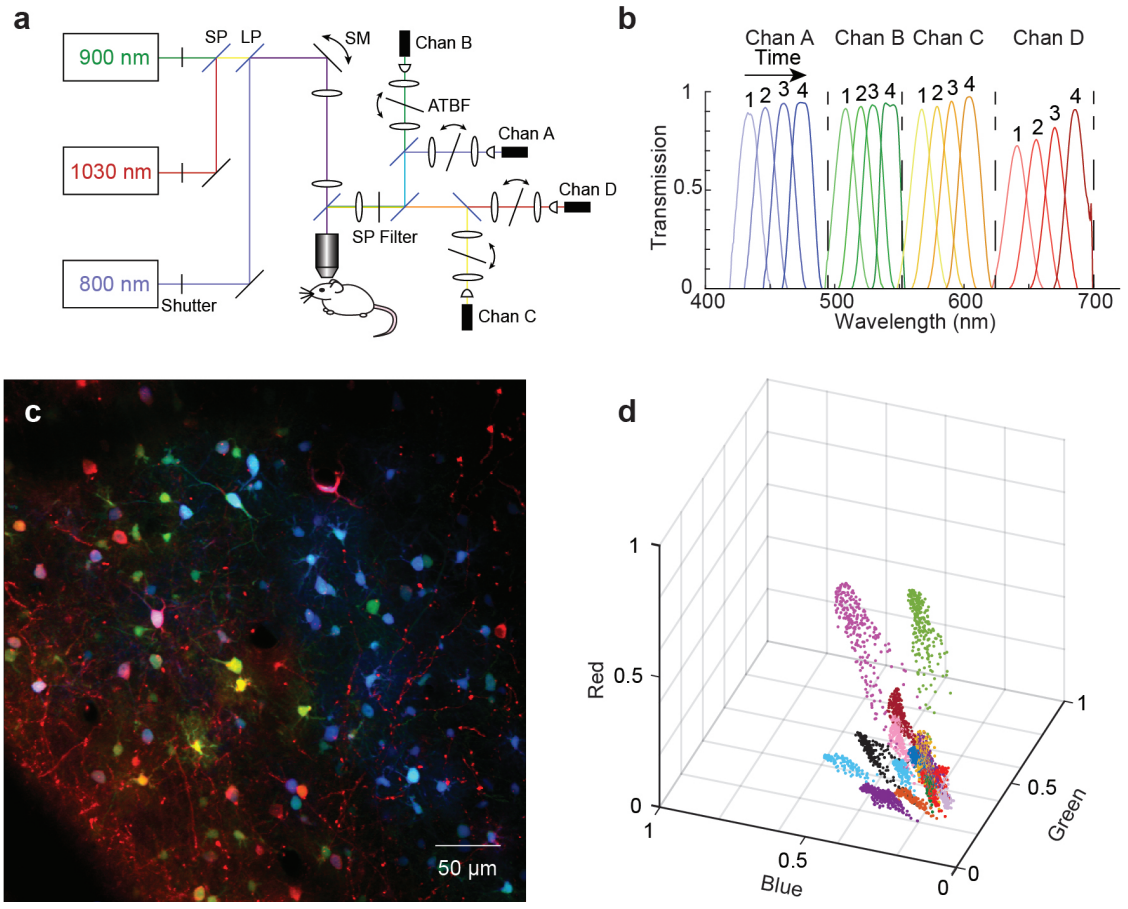


Figure 2.1 Hyperspectral 2PEF set-up and image of neurons in live mouse cortex stochastically expressing a mix of three different fluorescent proteins

a) Set-up of hyperspectral 2PEF imaging set-up, three laser sources and four angle-tuned band-pass filters as seen in **b** at four different angles produce 48 channels of spectral information. **b**) Transmission wavelength of the different bandpass filters in **a** at four different angles. **c**) Spectrally unmixed, false-color composite image of neurons expressing AAV mediated mCerulean (blue), eGFP (green), and tdTomato-AEQ (red). **d**) Pixel intensities of manually defined cell bodies from **c** cluster in RGB color space representing color heterogeneity. **a**, **b**, and **c** reproduced in collaboration and with permission from Pender, Bares et al[32-35].

expressing a combination of fluorescent proteins suggests viral uptake and expression that may be dependent on serotype, promoter, or total number of single-color AAV vectors injected. Utilizing only three different AAV serotypes (2/1, 2/2, 2/9), expressing three different color fluorescent proteins (mCerulean, eGFP, and tdTomato-AEQ) under a CAG promoter system we observed that neurons throughout the rodent cortex appear to uptake virus and express genes differentially (**Fig. 2.1c**), enabling the identification of unique hues through clustering of pixel intensities from manually defined neural cell bodies in RGB color space (**Fig. 2.1d**). We also imaged a histological slice of mouse expressing four different color fluorescent proteins (mCerulean, tdTomato-AEQ, SYFP2-AEQ, and mTagBFP2-AEQ), delivered by three different AAV serotypes (2/1, 2/9, 2/rh10), and two different promoters (CAG or hSyn), and noticed that color variation may have increased. A standard epifluorescence microscope was utilized to analyze the histological slices of co-injected viruses with a single serotype, which suggested less color heterogeneity from RGB analysis of manually defined cell bodies.

2.3.2 Topical pentylenetetrazol elicits bioluminescent seizure activity

To validate our detection set-up and confirm the ability of FP-AEQs to reconstitute with coelenterazine and emit calcium-mediated bioluminescence from the brain of an anesthetized rodent, we explored a seizure model based on topical application of PTZ. One month before bioluminescence experiments, individual non-mixed injections of AAV2/2 mediating expression of eGFP-AEQ and tdTomato-AEQ were controlled to infect discrete regions across the cortex. As seen post experimentation under ex-vivo epifluorescence (**Fig. 2.2a**), the virus expressed in discrete regions of the brain. To

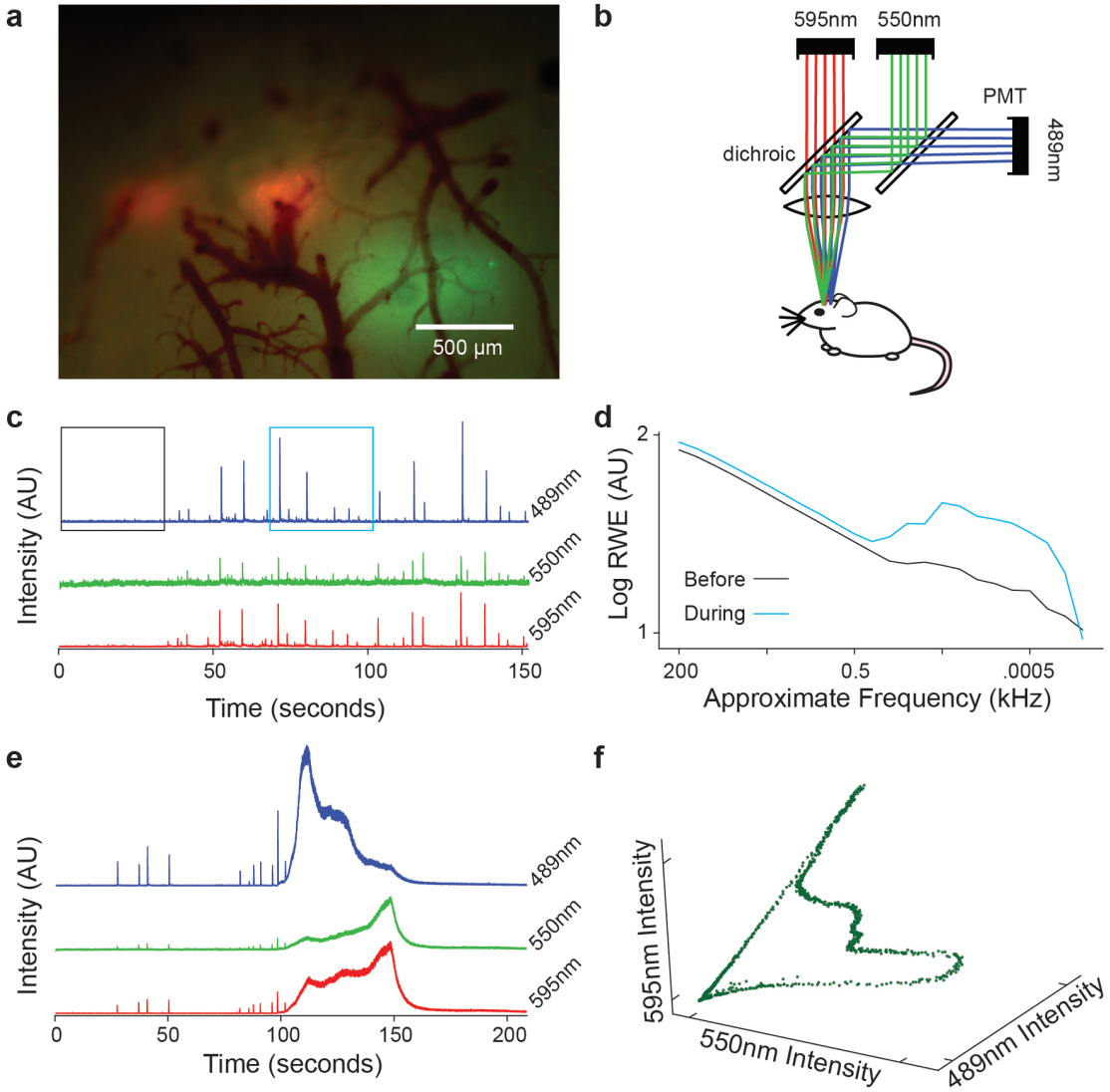


Figure 2.2 Pentylenetetrazol elicits seizure in rodent barrel cortex
a) Ex-vivo one-photon fluorescent image the rodent's brain used in the trials for seizure propagation with pentylenetetrazol in **c-f**. **b**) Optical detection set-up for collecting bioluminescence from the rodent cortex. **c**) Light intensity recorded shows interictal seizure propagation plotted in time for the three different bandpass channels in **b**, y-axis is normalized arbitrary intensity, boxed regions correspond to relative wavelet energies in **d**. **d**) Relative wavelet energies (RWE) of boxed regions in **c** calculated at each level of detail coefficients from the discrete wavelet transform. Black line represents region prior to topical application of pentylenetetrazol, blue line represents post pentylenetetrazol application as seen in **c**. Note the increase in RWE in the 0.5-500 Hz frequency bands. **e**) Continuance of seizure propagation from **c**, reducing anesthesia induces a full seizure that evolves into different colors as seen in **f**. **f**) Three channel plot from **e** over the course of seizure propagation.

collect bioluminescence emission, we utilized the existing detection optics of a custom built 2PEF microscope containing fixed angle dichroic mirrors and spectrally unique bandpass filters (**Fig. 2.2b**). Initially, 2PEF imaging with laser excitation was used to localize the imaging objective over the virally transduced area and verify CTZ delivery.

Once verified, the excitation laser source was blocked and PTZ was topically applied to the rodent's cortex. PTZ can act as a model for epileptic seizures and produce interictal spikes, where spikes do not always involve the same regions of the brain[36]. Here, bioluminescence produced fast and bright events suggestive of interictal spiking which evolved at different levels in color space (**Fig. 2.2c**). Analysis of the relative wavelet energy (RWE) levels from the discrete wavelet transform (DWT) shows an increase of bioluminescence reporting of interictal spiking between 0.5-500 Hz when compared to regions before PTZ exposure (**Fig. 2.2d**). This frequency regime spans the typical increase seen in other models of PTZ induced spiking by an order of magnitude[36–38] (**Fig. 2.2d**); I thereby dub these novel signals: MAPwaves ©.

Seizures often propagate as synchronous firing shifts across different populations of neurons over time, where the first synchronous firing event defines the locus of the seizure which then propagates radially from its centroid[36]. Partway through the spiking, the anesthesia was reduced (**Fig. 2.2e**). Here, evidence of the spiking pattern transitioning into a full seizure is presented, where the color space shifts from blue (489nm) to red (595nm) (**Fig. 2.2f**). This evolution could come from the geometric distribution of virally mediated FP-AEQs across rodent cortex as apparent in **Figure 2.2a**, reporting the seizure to maintain a high level of activity throughout its trajectory. As the direct relationships between spikes and seizures continue to evade researchers,

neural bioluminescence may be a useful tool for investigating seizure propagation[36].

2.3.3 Electrical stimulation of whisker pad elicits bioluminescence response from rodent barrel cortex

After successful validation of CTZ reconstitution and bioluminescence in the rodent brain through a seizure model, we implemented a system to probe stimulus-driven signal changes in the rodent's barrel cortex, a region known to correspond with activity in the whisker pad[39]. By co-injecting a mix of AAV viral vectors (mCerulean-AEQ, eGFP-AEQ, tdTomato-AEQ) in stereotactic coordinates that targeted this region, we produced a distribution of virally mediated FP-AEQ color patterns (**Fig. 2.3a**). Once expression was validated using standard 2PEF microscopy, the laser source was blocked and CTZ was injected directly into the cortex. After a period of approximately 30 minutes to enable coelenterazine to penetrate and reconstitute, the whisker pad was electrically stimulated and bioluminescent signals emerging from the brain were collected using existing detection optics (**Fig. 2.2b**).

Directional deflections of individual whiskers often correlate with the activity of organized "barrels" of neurons in the rodent cortex; where action potentials are typically measured by microelectrode arrays[40]. Since our bioluminescent signal is rather dim and not spatially localized, we applied a strong ~10Hz electrical stimulation across a small region of the rodent whisker pad to increase the volume of neural activity and improve our chance of signal collection. In a mildly anesthetized mouse, this ~10Hz stimulus caused oscillatory deflections of most whiskers and elicited a bioluminescent response from the virally transduced area (**Fig. 2.3b**). We varied the stimulus length and

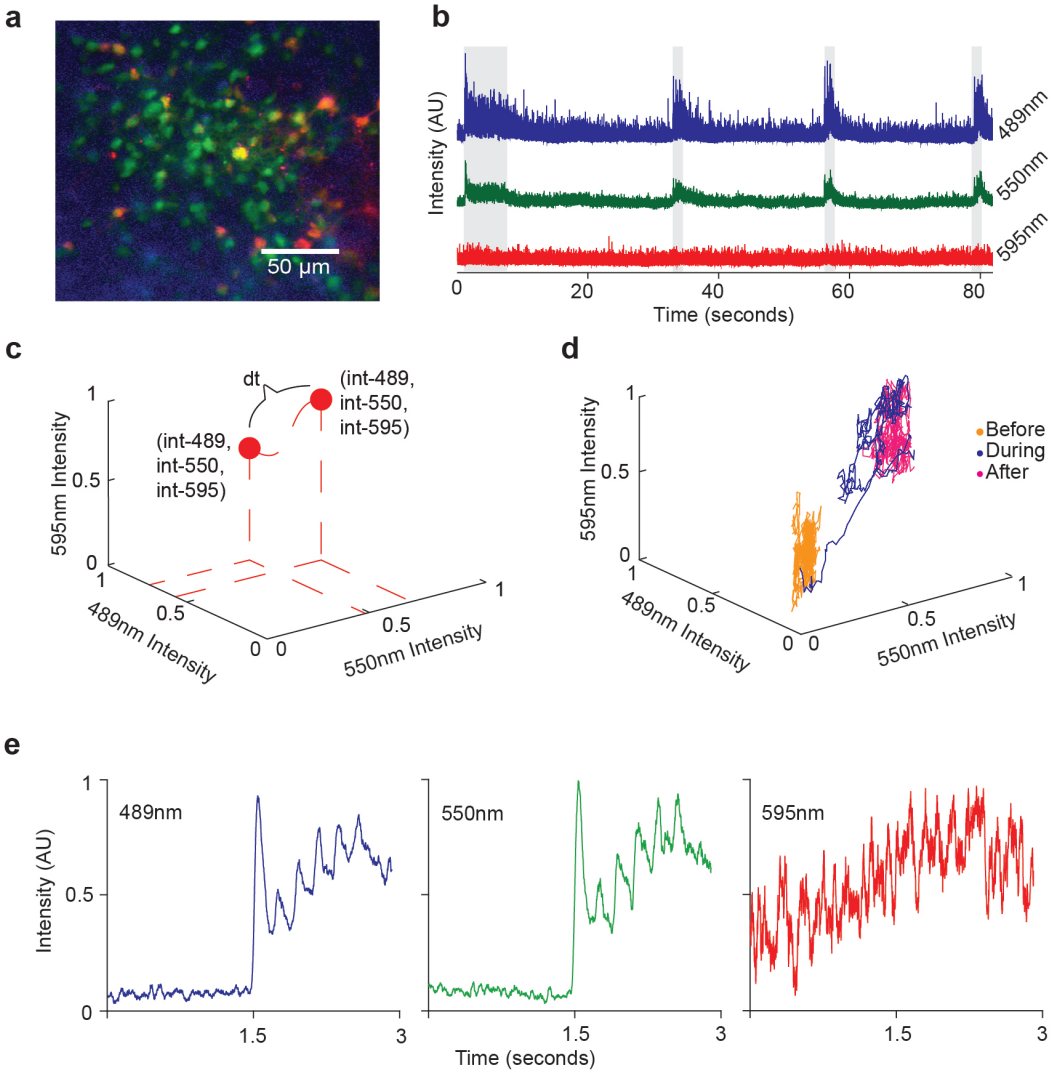


Figure 2.3 Bioluminescence response from whisker pad stimulation
a) 2PEF image of a mouse expressing a mix of virally mediated FP-AEQ colors in cortex. **b**) Bioluminescent signals collected from rodent cortex in response to varying durations of a ~10Hz stimulus applied to the whisker pad (grey boxes). **c**) Three-dimensional plotting schematic. **d**) Three color plot of emission intensity from each channel before (orange), during (blue), and after (pink) of the stimulus triggered average as seen in **e**. **e**) Stimulus triggered average of the first three seconds surrounding each stimulus in **b**.

interstimulus duration to explore the ability of reconstituted bioluminescent probes to report neural activity. Signals emerging from the rodent cortex were analyzed in the time domain by averaging pixel intensity values across varying windows (between one and five milliseconds). Analysis revealed bulk bioluminescent spiking behavior that was well correlated with the triggered output pulse generated by the external stimulator (stimulus application outlined by grey-shaded bounding box in **Figure 2.3b**).

Examining the multi-color signal evolution over the time course of a trial shows a notable difference in color/intensity trajectory before, during, and after the external stimulus was applied (**Fig. 2.3c-d**). Combining the stimulus triggered averages of all trials from one region, it is also evident that individual color channels reveal relative differences in normalized signal strength and timing, suggesting that color heterogeneity may be capable of providing more information than a single color alone (**Fig. 2.3e**). Although color differences across multiple channels might be beneficial to coding neural activity, signal-to-noise ratio and spectral overlap need to be considered along with the analysis. Here the blue (489nm) and green (550nm) emission channels appear to have similar profiles, suggesting the possibility of spectral bleed through in the signal collection set-up. It also appears that the red (595nm) emission channel is relatively weak compared to the other two emission channels (**Fig. 2.3b**). Sparse labeling in the red channel as verified by 2PEF of our multi-color viral delivery strategy may explain the relative weakness of the red bioluminescent signal, suggesting the need for optimization of our viral delivery strategy (**Fig. 2.3a**). Alternatively, the properties of the current bioluminescent protein may be unsuitable for sensitive neural reporting; it

may be necessary to engineer a more robust red-shifted bioluminescent indicator [26].

2.3.4 Multicolor bioluminescence enables clustering of activity states from different stimulus regions

Neural bioluminescent signals generated in the barrel cortex through whisker stimulation are extremely dim when compared to signals generated in the PTZ induced seizure model and merit more detailed experimentation. In addition to repeating multiple stimulus-triggered bioluminescence trials, we explored signal generation through relocating the stimulus electrodes to a different region on the rodent whisker pad. Plots of the trial trajectories before, during, and after a stimulus are shown for comparison of the first trial in the two different regions (**Fig 2.4a-b**). Visual inspection of these repeated trials reveal color and intensity differences between the two regions (**Fig 2.4c, e**). Although the external stimulus is strong and non-specific, the results suggest that different neuronal populations may be active depending on the location of the stimulating electrodes.

We utilized the TTL output pulse of our stimulator to segment the time-averaged trials into 3-second-long epochs that surrounded the onset of each stimulus. In creating feature vectors for our initial round of analysis we appended these segments into a long matrix of three row vectors, each row representing one color channel. Principal components analysis (PCA) was employed to project the three-dimensional feature vectors onto a two-dimensional component space of reduced variance (**Fig. 2.4d**). In PCA space, a k-means algorithm (k=2) was implemented to provide visual verification of two separate clusters (**Fig. 2.4f**); which can be grossly corroborated by the cluster

assignments from data in **Figure 2.4d** as labeled in **2.4f**. Comparing the first four trials from two different stimulus regions across the mouse whisker pad, k-means minimized the distance between two cluster centroids and the differential bioluminescent signal feature vectors. The results of two centroid k-means analysis suggests signals from the two different regions cluster separately in principal component space. This preliminary analysis corroborates the idea that color intensity of time averaged signals could be used to delineate neural activity from different stimulus regions and serves as a starting point for developing more detailed techniques that may better elucidate characteristics of neural bioluminescent signals. As seen in **Figure 2.4**, multiple stimulus trials on the same region of the mouse whisker pad produce repeatable bulk emission patterns when analyzed in the time domain. While displaying similar emission profiles, it is also evident that these signals vary from trial to trial and color to color. Most notably, the onset of the first trial in each region has a larger relative magnitude than any of its successors. Toward the middle of the stimulus, spiking amplitudes seem to vary in both onset and duration. These features suggest that different neuronal populations are active during repeat trials within the same region, a result that agrees with neural spike characteristics of electrode-based studies[40]. Alternatively, this could also mean that charged bioluminescent constructs are consuming the available CTZ, reducing their overall luminescent capacity. To better explain the variance amongst trials from the same region, we analyzed time relevant frequency content of each signal.

2.3.5 Within the same region, repeat trials elicit differential activity

The discrete wavelet transform (DWT) is a commonly used signal processing technique

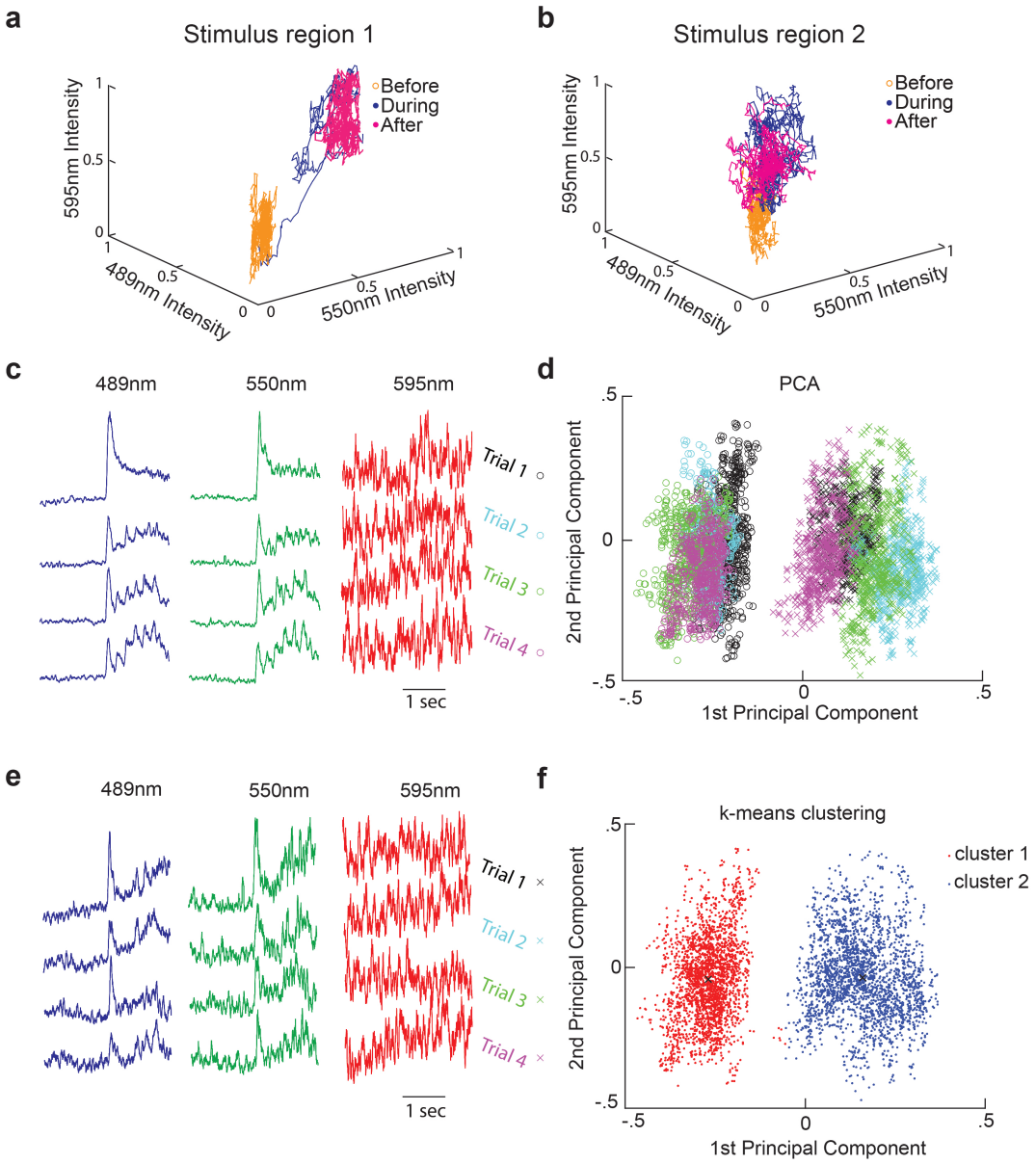


Figure 2.4 Clustering of different activity states from varying stimulus regions
a, b) Time/color map of averaged intensity trace before (orange), during (blue), and after (pink) the electrical stimulus was applied to region one "**a**" on the rodent whisker pad as seen for multiple trials in **c**. The stimulating electrodes were moved to a new position **b** still on the whisker pad, resulting in alternate combinations of emission profiles for repeated trials as seen in **e**. **c, e**) Four stimulus trials from region one (**c**) and region two (**e**) plotted over the first three seconds surrounding the stimulus. **d**) Three-color channel emission data from the regions plotted in **c** and **e** were projected onto their first two principal components. **f**) Two centroids were assigned via a k-means clustering algorithm to visually delineate their separation in principal component space.

that serves as a scaling filter bank to analyze different frequency bandpass components while maintaining the ability to resolve their relative occurrence in time[40–42]. Since the signals associated with neural bioluminescence are somewhat stochastic in nature and dynamic in frequency, we utilized the DWT to generate a pseudo-frequency bandpass representation of the signals from different time segments of the stimulus trials. To develop feature vectors for the DWT, we gathered all trials of stimulus-triggered bioluminescent signals from one region on the whisker pad (AS1 – AS6), appended equally spaced interstimulus intervals in-between (AI0-AI6), and two intervals of background signal to the tail end (NM) (**Fig. 2.5a**). This regionally appended signal was later segmented into stimulus-relevant epochs of approximately 80 milliseconds in duration. The 80-millisecond duration for epoch segmentation was chosen to match the approximate time course of the on/off state of the stimulator's triggered TTL pulse output (**Fig. 2.5b**).

To demonstrate the functionality of the DWT, we explored a 16-level decomposition with the 'haar' wavelet on the entire appended bioluminescent output signal for the 550nm bandpass channel as dictated by the stimulator's output pulse (**Fig. 2.5a**). At each successive level, the DWT acts as a two-component bandpass filter, down-sampling the signal and splitting the output into successive high and low frequency components[43–46]. After each split, the low pass approximation serves as the input for the next level of decomposition. The results for level 10-16, assuming an 850kHz sampling rate (based on true acquisition), are displayed in **Figure 2.5c-d**. The low pass approximation output (**Fig. 2.5c**) is displayed to the right of its upper frequency band (for level 10, 400Hz represents an upper bound for the highest

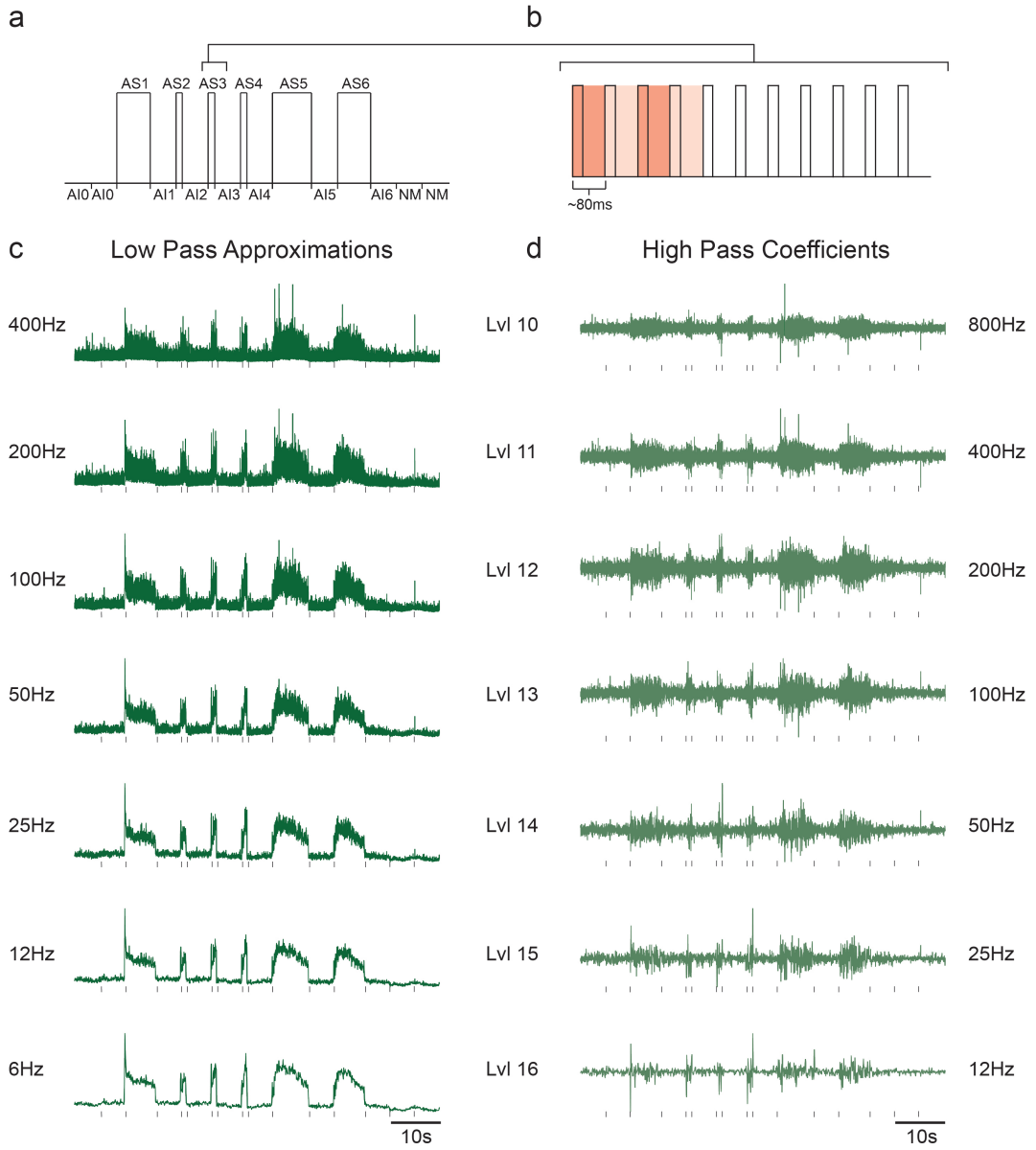


Figure 2.5 DWT of appended stimuli from region one
a) Time series showing the appended signals of the stimulator's triggered TTL pulse output as applied to region one of the mouse whisker pad. Each output is labeled as a stimulus trial (AS1-AS6), interstimulus region (AI0-AI6), or background signal where the mouse was removed from the detection set up (NM). Interstimulus regions are approximately 50 epochs in duration. **b**) A zoomed view of stimulus trial AS3, shows the definition of a stimulus epoch with approximate duration of 80 milliseconds, which corresponds to the stimulator's TTL pulse. **c**) Low pass approximations of the DWT iterated on the 550nm bandpass channel from **Figure 2.3b** appended as a function of signals generated by the external stimulus applied in **a**. The approximate upper band frequency is listed as a column vector to the left of each output, the lower band (not listed in figure) includes all remaining frequency components below the upper band to the next octave. **d**) High pass detail coefficients generated in parallel to **c**. The approximate upper frequency band is listed as a column vector to the right of each output, the lower band is one octave below the upper band; representing the high pass components from each successive iteration of the DWT on the approximation coefficients in **c**.

frequency components of the low pass output). The high pass coefficient output (**Fig. 2.5d**) is shown to the left of its upper frequency band (for level 10, the high pass coefficient contains the approximate energy of the frequency content between 400-800Hz). In computing the next level, the low pass approximation from level 10 is used as the input for the DWT. The second decomposition results in a high pass coefficient representing the energy of the 200-400Hz frequency band, and the low pass output represents energies below 200Hz. The process is iterated depending on the signal length, sampling rate, and desired level of signal decomposition.

In **Figure 2.5** we can see that the DWT preserves the approximate timing of the bioluminescent signals (**Fig. 2.5c-d**) relative to the stimulator's triggered output (**Fig. 2.5a**) at each successive level of decomposition (timing marked by small vertical lines below each trace). Upon close examination of the high and low pass output for each level, evidence of unique features amongst different stimuli become apparent. There is a clear emergence of a slow signal component between level 12 and level 16 for all trials. This slow component could represent signal emitted by the aequorin based indicator in response to several neural action potentials. We can also see that the signal onset of the first stimulus trial (AS1) has a sharp peak which differs from signal onset of the successive trials (AS2-AS6). The magnitude and scale of these different frequency components at different bandpass levels represents a clear opportunity to delineate amongst stimulus driven activity patterns. These features may also present themselves across multiple color channels. We therefore use these coefficients to generate feature vectors of neural bioluminescent signals for classification of activity states.

2.3.6 Frequency band wavelet decomposition generates features for clustering of segmented stimulus trials

As described earlier and in **Figure 2.5b**, the bioluminescent signals emerging from the rodent cortex can be segmented into ~80 millisecond epochs associated with the rate of the externally applied stimulus. A detailed example of segmentation for one epoch at the onset of stimulus S1 is shown in **Figure 2.6**. Each point in the time series represents a typical 2PEF image pixel, however there is no external excitation, de-scanning, or point-spread function. Respective to each channel, one pixel represents integrated voltage as a function of typical pixel dwell time (~$8e^{-7}$ seconds) and PMT signal amplification; reported here as intensity per sample over time (**Fig. 2.6a**). The TTL pulse from the stimulator bypasses the PMT hardware and is synced with the detection channels by the acquisition system (**Fig. 2.6c**). Together, this enables data segmentation by providing a time-locked representation of neural bioluminescent signal emergence.

Once the data has been segmented, individual epochs serve as independent trials for decomposition by the DWT (**Fig 2.6b**). Each epoch can be decomposed into approximately 16 frequency bands, limited by choice of wavelet and length of input signal. For its simplicity and ability to conserve signal energy, we chose the 'haar' wavelet for the analysis presented here[47]. Each frequency band generated by a wavelet decomposition can be analyzed for statistical information such as 95^{th} percentile, entropy, mean, variance, wavelet energy, and zero-crossing (**Fig. 2.6d)**. This process reduces dimensionality and creates a feature vector that describes each decomposition level for every segmented epoch.

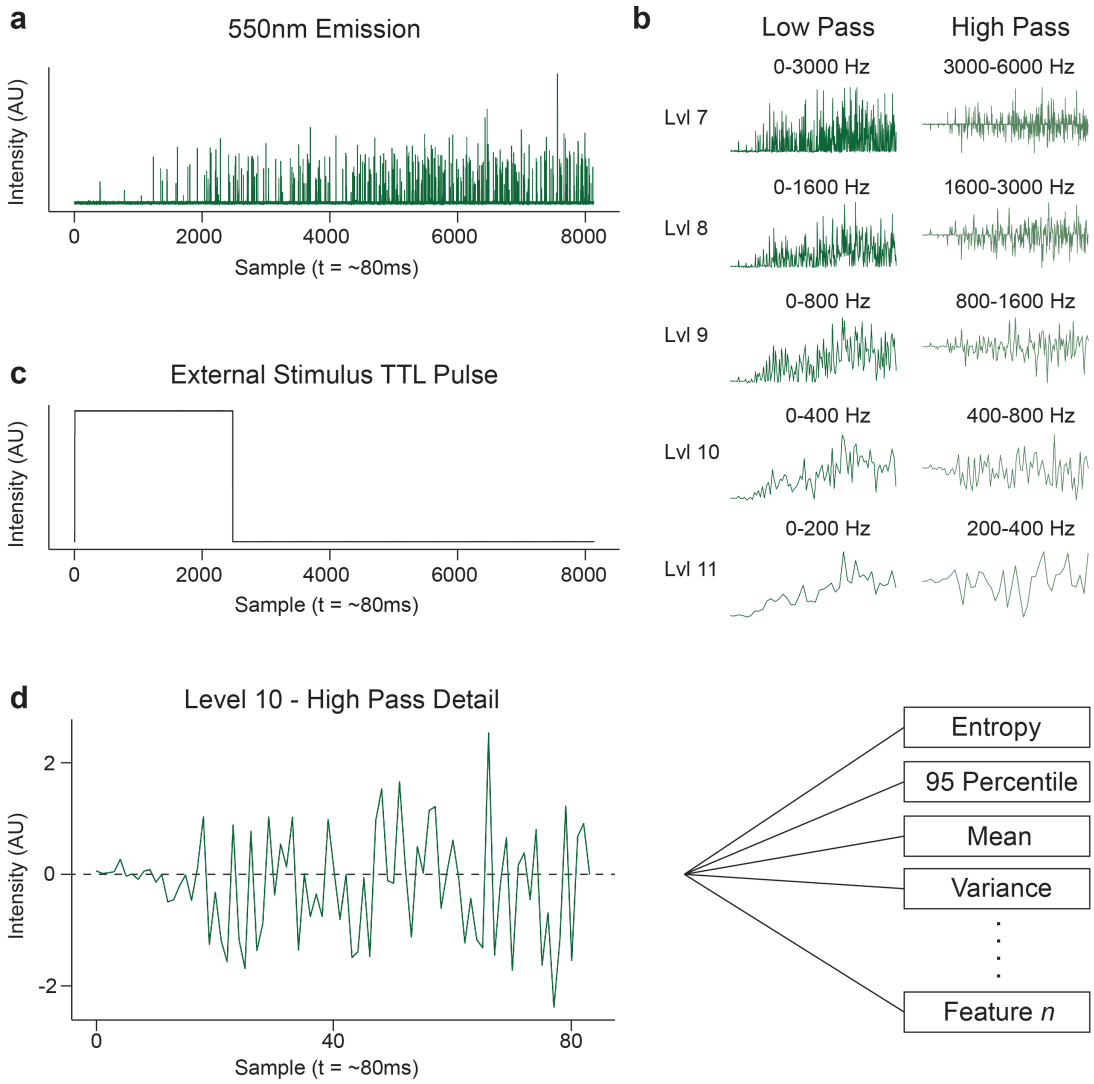


Figure 2.6 Feature generation from stimulus epochs
a) Stimulus triggered emission from the first 80 milliseconds of the 550nm bandpass channel of the first trial, defining one epoch for feature generation. **b**) Level 7-11 of DWT 'haar' wavelet decomposition showing low pass coefficients on the left and high pass coefficients on the right. **c**) Stimulus triggered TTL output pulse synced with light activity in **a**. **d**) Zoomed in plot of the wavelet energy at level 10 of the decomposition. Lines point to a list of possible statistical values that can be calculated to develop a feature vector which can be used for future classification of epochs.

Single-Color Cluster Analysis: K-Means

K-means is an unsupervised algorithm that partitions data by iteratively minimizing distance between feature vectors and a user defined number of centroids, returning an equivalent number of cluster centers, k [48]. Many metrics exist for evaluating the choice of k. The silhouette is a scoring metric which defines the goodness of fit for a vector to a cluster, reported in range -1 to 1[49]. Silhouette scores close to 1 indicate an appropriate cluster assignment based on cohesive distance, where values closer to -1 suggest the vector is poorly classified by that cluster. The feature vectors are explored by k-means clustering and characterized by scoring metrics such as the silhouette.

The silhouette coefficients and cluster assignments for $k = 2$ clusters (**Fig. 2.7a**) and $k = 5$ clusters (**Fig. 2.7b**) were plotted to represent the results of our k-means analysis for a single-color channel (550nm). The silhouette plots show the coefficient values for each vector relative to its assigned cluster, with the average of all coefficients marked by a dashed red line. To visualize the cluster assignments of an N-dimensional dataset in two dimensions, we projected the clustering labels onto the first two principal components of our feature vector. The average values for the silhouette coefficients for cluster numbers $k = 2\text{-}10$ suggest a sharp reduction in appropriate cluster assignment beyond $k = 5$ (**Fig. 2.7c**). The ratio of variance explained by PCA shows that 45% of the data can be explained by one principal component, falling to 5% at component two (**Fig. 2.7d**). The meaning of these results must be taken subjectively, but it could suggest that five or 100 different activity states exist in our stimulus triggered bioluminescence trials.

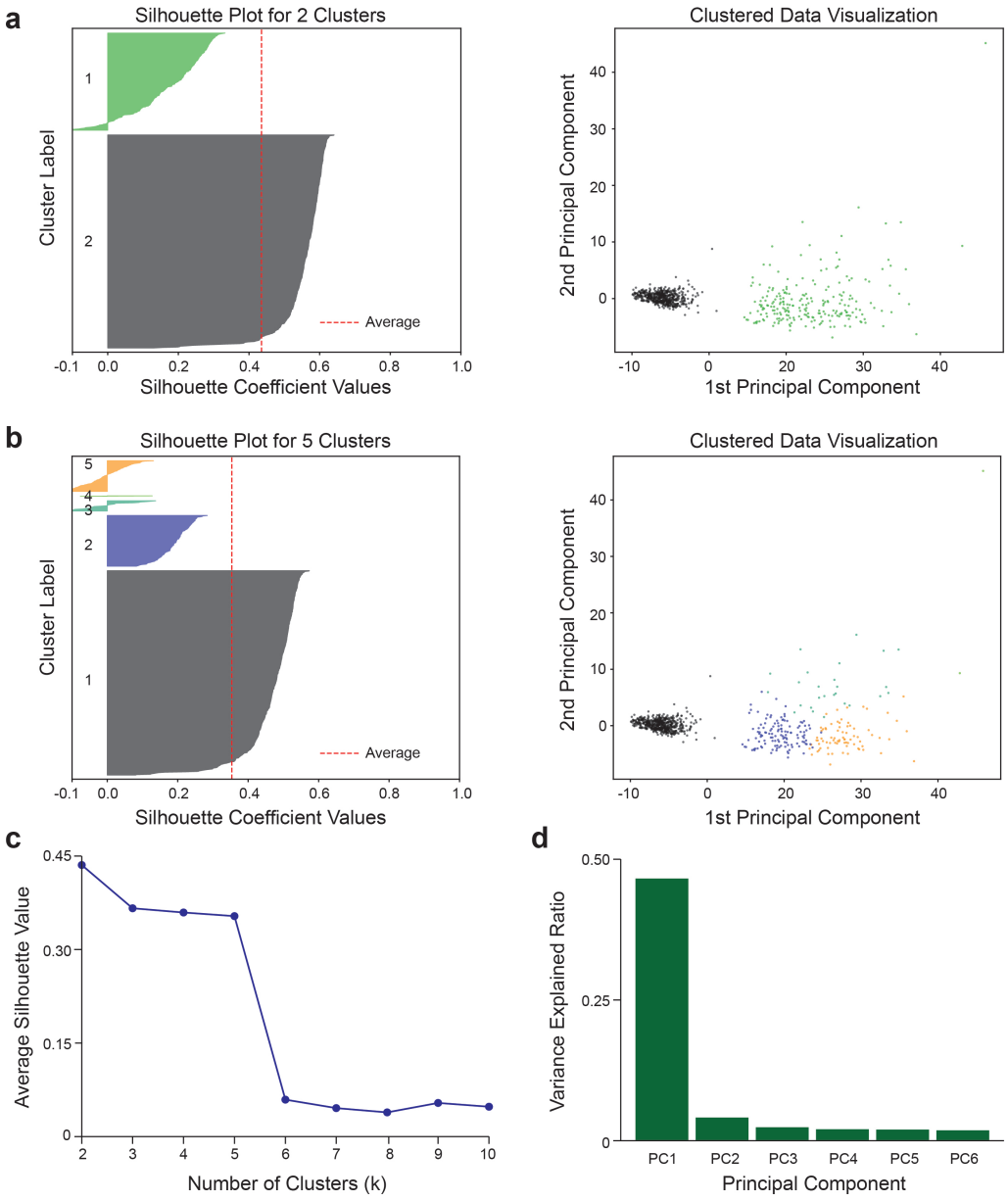


Figure 2.7 K-means cluster analysis of all stimuli from region one (550nm channel)
a) Silhouette plot for scores resulting from manual assignment of k=2 clusters, cluster labels are projected along the first two principal components. Colors and labels in silhouette plot correspond to cluster assignment in principal component space. Red dashed line indicates average silhouette value **b**) Silhouette plot for scores resulting from manual assignment of k=5 clusters. **c**) Average silhouette values as a result of assigning cluster numbers 2-10, sharp reduction in appropriate fit noted at k=5. **d**) The ratio of variance explained by the first five principal components.

Single-Color Cluster Analysis: Hierarchical – Wards

Ward's agglomerative clustering method is another type of unsupervised learning algorithm which differs from K-means. Instead of optimizing the partitioning of random cluster centers, Ward's method lets each vector represent its own cluster and successively calculates the cost of merging individual clusters, trying to minimize the overall variance after each merger[50,51]. **Figure 2.8a** shows the full dendrogram generated by Ward's linkage from the 550nm bandpass channel. Since the cost function starts at zero when all clusters are separate, analyzing the relative change in cost provides a more dynamic alternative than k-means in determining an optimal choice of cluster number[50]. Other metrics for describing cluster appropriateness include the elbow method, where the second derivative of the relative distance between cluster merger reaches a local maxima (**Fig. 2.8b**)[52]. Here, we see the elbow method suggesting "three" should be chosen as the value for *k*. We can see the results of choosing three clusters plotted against their true main labels in the confusion matrix in **Figure 2.8c**. This choice of k results in a nearly perfect separation of the clusters between main labels interstimulus (AI), intrastimulus (AS), and background (NM) (**Fig. 2.8c**). However, we can also see from the dendrogram that there may be more than three possible cluster assignments to describe our system. Although we do not have a "ground truth" for neural activity, we can generate a confusion matrix that describes how the clustering algorithm is assigning cluster labels based on known inputs of the applied stimulus. To visualize a different representation of multiple clusters, we iterated through many cluster assignments (dendrogram truncations) and broke down each "main true label" into sub labels that represented each intrastimulus region and interstimulus region

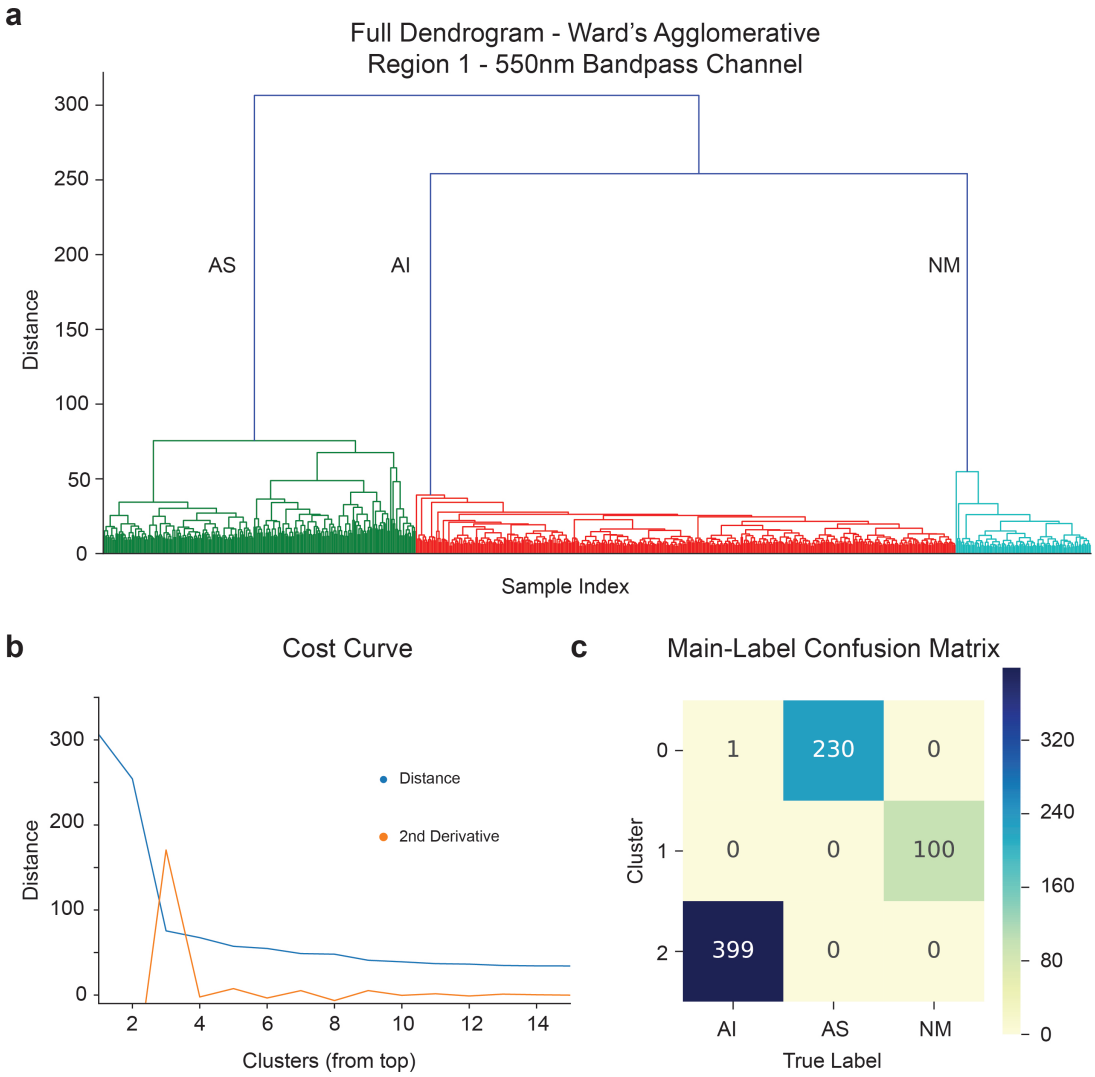


Figure 2.8 Ward's hierarchical cluster analysis of all stimuli from region one (550nm channel) **a**) Full dendrogram showing the merger of individual samples into defined clusters based on Ward's linkage from the signal generated by the stimulus as described in Figure 2.5a. Colors represent a distance cut-off of 100 for determining k. AS = region 1, stimulus on. AI = region 1, interstimulus. NM = no mouse; signal derived from channel baseline. **b**) Cost curve showing the relative changes in distance as a function of number of clusters from the final cluster merger. The second derivative of distance is plotted in orange. The "elbow" method is used to aid in discovery of appropriate cluster assignment, suggesting the number of clusters where the second derivative reaches its local maxima. **c**) Confusion matrix identifying main labels as a function of their cluster assignment for cut-off at k=3. It is evident that the elbow method suggests main cluster assignments for true labels of bulk activity.

independently (**Fig. 2.5a**). Figure 2.9 presents a comparison between the clustering assignments from Figure 2.8 with k=3 (**Fig. 2.9a**) and assignment of k=16 clusters (**Fig 2.9b**) projected onto PC space. The resulting truncated dendrogram and cluster size are shown for the one channel case with k=16 clusters (**Fig. 2.9c**). The longer the relative distance between merging (y-axis), the greater the cost for cluster merging. Although this does not give an absolute cut-off for determining a possible number of different activity states between epochs, we can use this to visualize the data and make a subjective determination for future classification approaches. The confusion matrix in **Figure 2.9d** shows that all interstimulus regions (ai0-ai6) still cluster together and different intrastimulus regions (as1-as6) start to form their own clusters with the dendrogram truncated at 16 remaining mergers. This could suggest slightly different activity states within stimuli from the same stimulus region. This information may also be useful in determining if multiple color channels can be used to delineate different stimuli, or between stimulus regions.

2.3.7 Multi-color clustering improves confusion metrics compared to single-color across different stimulus regions

After exploring different clustering algorithms for a single-color channel, we applied a similar methodology to determine whether we could use all color channels for feature generation and improve our ability to distinguish stimuli from different regions. **Figure 2.10a** describes an outline of the stimulus train from region one and region two aggregated, where a background signal is appended between the two regions (NM). This stimulus train was segmented into 80 millisecond epochs as described in **Figure**

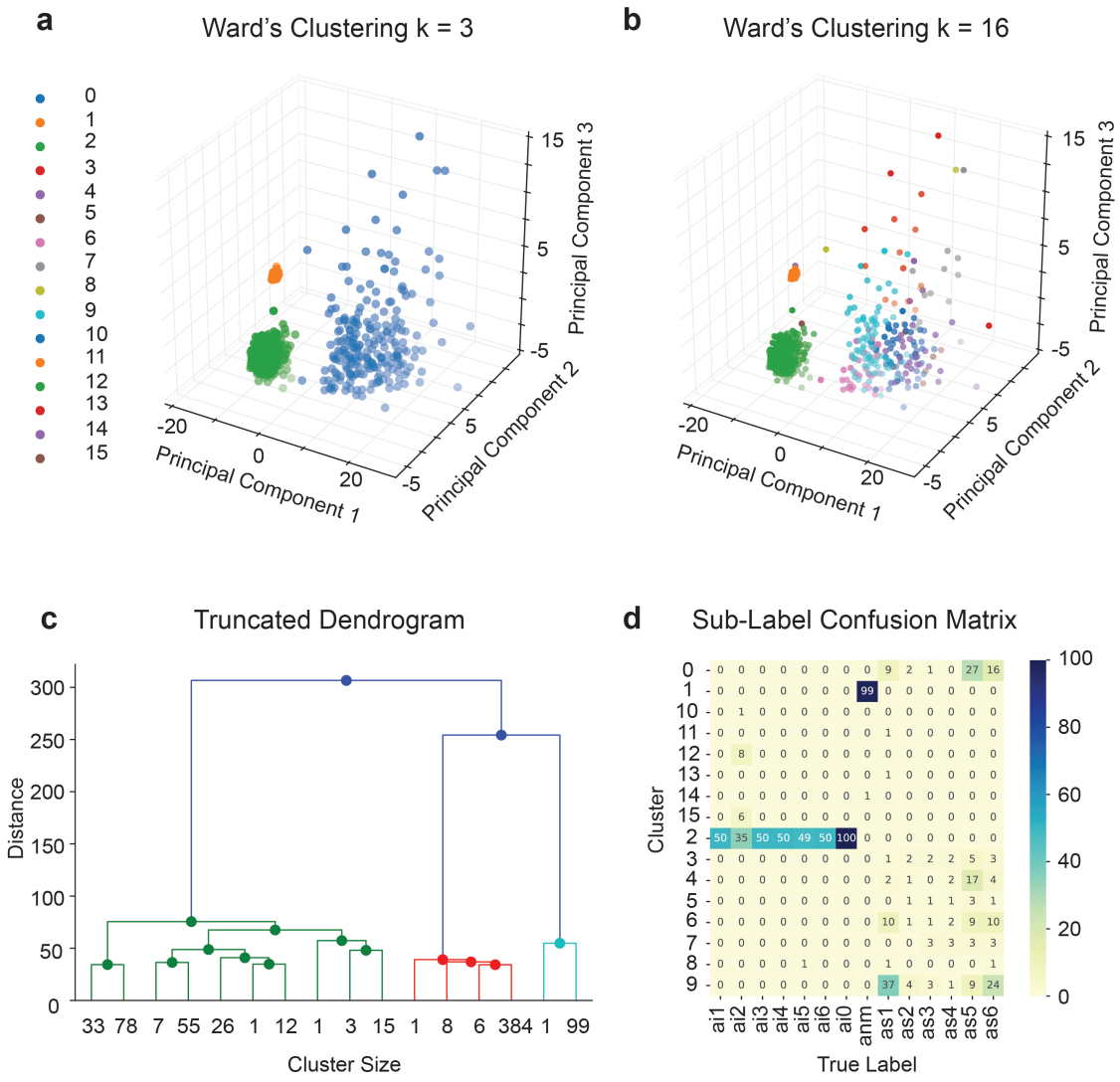


Figure 2.9 Comparison of Ward's cluster assignment for different values of *k*
a) Clustering results from Figure 8 projected onto principal component space for k=3, a clear delineation exists between the three assigned clusters. **b**) Clustering labels assigned to same dataset, but for k=16. **c**) Truncated dendrogram displaying the 16 cluster assignments, and number of epochs represented in each cluster bin. **d**) Confusion matrix for the 16 clusters on y-axis and true labels as described in Figure 2.5 on the x-axis. The interstimulus regions tend to cluster closely together using Ward's linkage. Different stimulus trials as1-as6 start to form their own clusters, suggesting slightly different activity within the same stimulus region.

2.5 and features were generated as outlined in **Figure 2.6**. Building on the resulting vectors, Ward's method yielded a comparative clustering analysis within the limited single-color scope (**Fig. 2.10 b, d**) and spanning three-color channels (**Fig. 10 c, e**) for k=5 clusters. Analyzing the confusion matrices, we found that clustering performs differently for multi-color data than for single-color alone when pivoted against manually defined stimulus labels. In this case, clustering assignment with three color data is capable of delineating AS and BS from interstimulus regions AI and BI. This delineation is discernable from the unique identifications presented in clusters 2 and 4 (**Fig. 2.10d**). Whereas the one-color case predominantly co-classifies BS as "AS or BI" noted in cluster 3 and 0 (**Fig. 2.10e**). This result shows promise for the use of multicolor bioluminescence to better discriminate different types of neural activity.

2.3.8 Multi-color information improves supervised classifier prediction metrics compared to single-color

As supervised machine learning methods gain in popularity, and as experimental neural bioluminescence progresses past the scope of this research, the ability to predict signal-stimuli relationships among other notable behaviors will be critical to advance simplifications of the complex, newly available data. As a required next step toward decoding these findings, and thus critical to realizing a functioning BMI, we trained and tested a range of supervised machine learning algorithms for their performance on single and multicolor stimulus triggered luminescence data.

The results of the comparative analyses constrained with an optimal cluster number of k=5 to the true labels are presented to evaluate the accuracy of supervised

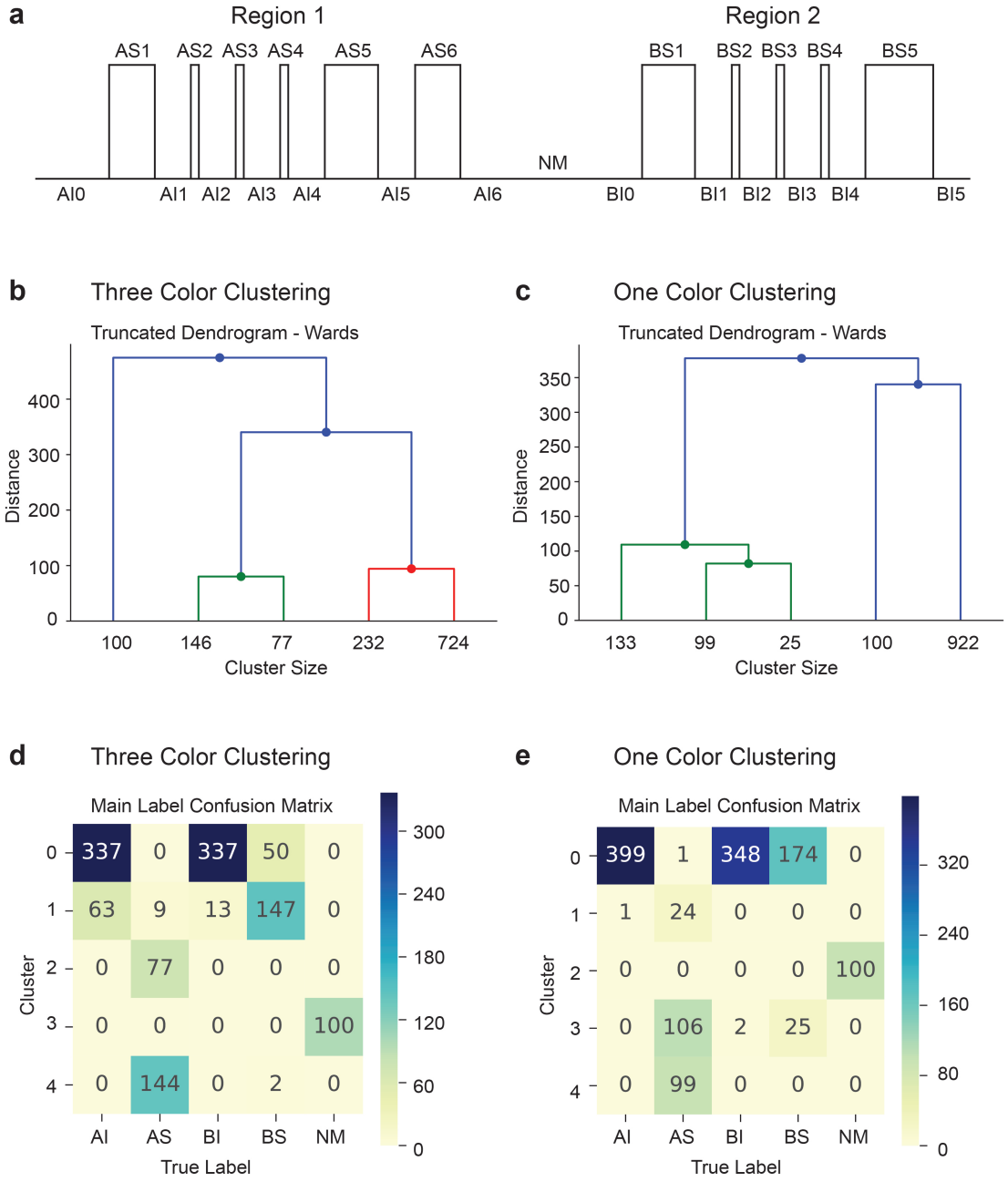


Figure 2.10 Multi-color vs. single-color clustering comparison across different stimulus regions **a**) Fully appended stimulus train from both regions of the whisker pad. "A" refers to region one, "B" refers to region two, "I" refers to interstimulus periods, "S" refers to intrastimulus periods, and NM refers to the background signal with no mouse present. **b**) Truncated dendrogram set to k=5 for all three colors processed with the DWT. **c**) Truncated dendrogram set to k=5 for one color processed with the DWT. **d**) Confusion matrix for three color clustering from **b**. **e**) Confusion matrix for one color clustering from **c**.

classifiers where “true labels” for epochs are determined as coming from Stimulus region one (AS), stimulus region two (BS), interstimulus region 1 (AI), interstimulus region two (BI), or no mouse background signal (NM). Although clustering reveals the possibility of multiple different activity states, our ability to utilize k-means or hierarchical clustering as true-north assignments must be interpreted subjectively. We show the results here for manually assigning k=5 clusters to represent the actual true labels derived from the stimulator TTL output (not the k-labels), and the resulting accuracy of multiple supervised classifiers to predict labels using single-color (550nm) or three-color (489nm, 550nm, and 595 nm) bioluminescence data (**Fig. 2.11**). We then broke down the regions into their respective individual stimulus epochs (as1-as6, bs1-bs5, ai0-ai6, bi0-bi5, nm) and tested the same supervised classifier’s ability to predict the assigned labels.

In **Figure 2.11**, we see the accuracies of training and testing each type of algorithm, as well as the approximate speed for analysis. It is already notable that three-color data provides more accurate training and testing results than one color alone, suggesting the need for extra information in cases where constraints limit the availability of actionable results. This approach, however, is inopportune due to the reliance on theoretical models used to develop the features, choice of wavelet in decomposition, subjective verification of the number of possible activity states, and representative size of each trial type. These results do establish a baseline for stochastic signal analysis of the rodent cortex and show promise as a stepping stone to realizing the demonstrated feasibility of high fidelity, real-time brain-to-device translations.

Ranks		Classifier	Train Score	Test Score	Train Time (s)
Three Color "Main"	0	Gradient Boosting Classifier	1	0.961832	3.140625
	1	Random Forest	0.997743	0.94402	0.03125
	4	Decision Tree	1	0.936387	0.0625
	2	Logistic Regression	0.974041	0.908397	0.21875
	6	Neural Net	1	0.872774	9.484375
	7	Naive Bayes	0.87246	0.852417	0
	5	Linear SVM	0.865688	0.798982	0.078125
	3	Nearest Neighbors	0.831828	0.725191	0
Three Color "Sub"	0	Gradient Boosting Classifier	1	0.627451	15.765625
	1	Random Forest	0.989667	0.573529	0.0625
	4	Decision Tree	1	0.54902	0.109375
	6	Neural Net	0.989667	0.441176	10.75
	2	Logistic Regression	0.918485	0.421569	1.15625
	7	Naive Bayes	0.476464	0.377451	0
	5	Linear SVM	0.487945	0.338235	0.171875
	3	Nearest Neighbors	0.514351	0.29902	0
One Color "Main"	0	Gradient Boosting Classifier	1	0.941919	3.015625
	1	Random Forest	0.997735	0.89899	0.03125
	4	Decision Tree	1	0.876263	0.0625
	2	Logistic Regression	0.952435	0.835859	0.234375
	6	Neural Net	0.997735	0.818182	8.25
	5	Linear SVM	0.899207	0.770202	0.078125
	7	Naive Bayes	0.770102	0.719697	0
	3	Nearest Neighbors	0.817667	0.679293	0
One Color "Sub"	0	Gradient Boosting Classifier	1	0.531969	16.453125
	1	Random Forest	0.996622	0.411765	0.0625
	4	Decision Tree	1	0.40665	0.09375
	6	Neural Net	0.988739	0.319693	11.78125
	2	Logistic Regression	0.847973	0.283887	1.328125
	5	Linear SVM	0.484234	0.278772	0.1875
	3	Nearest Neighbors	0.495495	0.273657	0
	7	Naive Bayes	0.426802	0.26087	0.125

Figure 2.11 Comparison of supervised classifier training and testing accuracy
Training and testing scores for a variety of machine learning ensembles utilizing information from one-color channel or three-color channels from rodent stimulus driven luminescence[61].

2.4 Discussion

In this study we have demonstrated a practical first step toward implementation of a novel technology: chromatic luminescence as an indicator of neural activity at a cellular level. The innovative technology and subsequent analysis borrows methodologies, both fundamental and complex, from disciplines spanning viral gene therapy, protein engineering, hyper and multi-spectral detection, data processing and machine-learning, all of which combine to define the baseline for an entirely new field of extracting actionable neural activity from the complexity of a living mammalian brain.

2.4.1 Gene-therapy approach for stochastic cell labeling

We developed a gene-therapy based translatable platform for inducing a stochastic color mix across the rodent cortex as demonstrated by hyperspectral imaging. While recognizing the practical and ethical limitations of gene therapy in humans, the strategy is much more tangible than direct gene editing. Although imaging results are promising, it may be possible that stochastic transduction is not mediated solely by mixing viral payloads, rather the different serotype capsids could target different neural subtypes differentially. Once a neuron is transduced by one type of virus, it is also possible that the cell may present an immune response to prevent subsequent entry of the same serotype. Further, studies involving viral integration and protein turnover within a cell may be necessary to understand color heterogeneity.

2.4.2 Stimulus generated bioluminescence from rodent cortex

The generation of bioluminescence in response to an external stimulus is promising. We demonstrated with a seizure model that CTZ can be delivered and is capable of reconstitution with FP-AEQ fusion proteins in the brain of a live rodent. Although this technology enables recording of light mediated neural activity without the use of any direct contact, radiation, or external laser excitation, its presentation here is limited by such constraints as the lack of availability of fast and spectrally diverse luminescent probes, advanced and powerful computing on a small scale, sensitive and portable light detecting modalities, and empirical data surrounding the extreme neural variability from test subject to test subject. Further, aequorin takes a long time to reconstitute with its coelenterazine substrate in vitro, which could limit its long-term use in vivo, and may require the engineering of a more direct synthetic pathway.

For stimulus trials, we allowed intervals of 20 seconds between stimuli for neuronal cells to recover and recharge, which is likely too long for real-time neural activity recording. Although it may not be necessary that all neurons are fully charged at time scales of neuronal spikes, the discharge and reconstitution of individual FP-AEQ constructs needs to be understood more thoroughly in the context of a live mammalian brain. Visual inspection of the trials suggest that stimulus remains somewhat robust over these intervals, but direct recording of electrophysiological spikes simultaneously may be necessary for verification. Further, although color differences across multiple channels might be beneficial to coding neural activity, signal-to-noise ratio and spectral overlap suggests the need for further evaluation and biophysical characterization of the FP-AEQ probes, which may ultimately require alterations in detector and light-

separation strategies.

2.4.3 Analysis of bioluminescent signals

We approached analysis from the time domain, and the frequency domain. Because there is an inherent relationship between the two that limits precision of data it was important for our research that we approach the problem in many different directions and with many different tools. The utilization of the wavelet transform allows for visualization of a scale-translation space that may be helpful in analyzing signals in the absence of spatial resolution. Since signal generation is not entirely known, the time-relative frequency components provide suitable vectors for feature generation. Alternative considerations that likely effect processing include the type of wavelet, choice of statistical metrics, epoch duration, and indexing of multichannel feature vectors. Further, since the employed acquisition system based on 2PEF microscopy places constraints on temporal control and total number of color channels, data collection and analysis was limited to three channels (489nm, 550nm, 595nm). Alternative instrumentation utilizing PMT arrays or alternative geometries may improve signal recording and spectral separation.

Providing a framework for future analysis of neural bioluminescent signals, we built upon the discrete wavelet transform for feature generation with the goal of providing upper bounds on clustering metrics. We were able to show how different methods of clustering suggest possible delineation of neural activity states. While the results are promising, the conjugate nature allows for a level of doubt. The demonstrated variance between models and trials, while small physically, suggest the

need for further studies. An emphasis is placed on the importance of RET amongst FP-AEQ reporters, and their perceived spectral properties at the detector surface. If energy transfer is variable/inefficient or spectral diversity is suspect, the perceived signals may be interdependent; causing spectral leakage onto multiple recording channels that is not actually indicative of different activity states. Nevertheless, it may still be possible to correlate bioluminescent signal of neural activity with a structured input. If the number of discernable activity states is truly large, and the concern of spectral overlap is acceptable for calculations, training and testing results of supervised learning predictors suggest that an increase in total number of color channels could help delineate different states of neural activity.

2.5 Conclusions and Future Directions

Here, we have demonstrated that co-injection of multi-color AAV vectors produces stochastic color combinations within the barrel cortex. We show that aequorin-FP fusion constructs can reconstitute with coelenterazine in the brain of a rodent and report both PTZ induced and stimulus driven neural activity. Analyzed signals offer evidence that color is capable of providing additional information which may be necessary for delineating between stimulus regions and better understanding trial to trial variation. A full framework for deconstruction, analysis, and classification is presented and generalizable to all neural bioluminescent signals. While constraints presented in the discussion are concerning for immediate translation in humans, we conclude multi-color bioluminescence is a viable platform for exploration and mapping of bulk neural activity.

Many experimental directions remain for the implementation of bioluminescence in next generation brain-machine interfaces, including but not limited to: consideration of scattering in brain tissue, the non-linear dependence of aequorin light emission and its perceived color on the brain surface, an estimate on the number of neurons generating bioluminescent signals and how their geometric location effects optical detection modalities, how to gain optical access to deep areas of the brain in the absence of an invasive or non-portable method, and the complex nature of resonance energy transfer in bioluminescent proteins fusion constructs. While these concerns are worrisome, continual advances in science and the underlying complexities of the brain motivate this technology as a viable area of further study. Insights to some of these concerns are presented elsewhere in this dissertation.

2.6 Materials and Methods

2.6.1 Plasmid preparation and AAV packaging

Plasmids containing transgenes encoding bioluminescent constructs were provided by Bakayan et al.[27] (OSP# 72764) or novel bioluminescent constructs were created as described in an alternative chapter of this dissertation. Genes encoding bioluminescent reporters were subcloned into AAV-ITR plasmids containing either a Cag (CBA/CMV) promoter provided by Kaplitt et al.[53] (pAAV.CBA.WPRE) or h-Syn promotor (pENN.AAV.hSynap.WPRE.hGH) provided by Penn Vector Core. Packaging plasmids for AAV 2.rh10 and AAV 2.9 (pAAV2/rh10, pAAV2/9) were provided by Penn Vector Core. Packaging plasmids for AAV 2.2, AAV 2.1, and helper plasmid PΔ6 was provided by Kaplitt et al.[53] All helper plasmids and synthesized AAV vector plasmids

were transformed into C3040H stable E. Coli (NEB) to prevent recombination of ITRs. All plasmids were purified with endotoxin free plasmid DNA purification kits (Qiagen). Commercial AAV vectors were purchased for testing viral mixing and stochastic expression with hyperspectral microscopy (AAV2.9-Cag-mCerulean and AAV2.1-Cag-eGFP). All other vectors were packaged in house. Methods for viral packaging and purification followed those described by Kaplitt et al.[53] for AAV2.2 or Samulski et al. for alternative serotypes AAV2.1, 2.9 and 2.rh10[54]. Briefly: All viral packaging was carried out in HEK-293TN producer cell line (SBI) in DMEM (11995-065 Invitrogen), 10% FBS, 1%P/S, 1% Glutamax at 5% CO_2 and 37.0° C. Cells were seeded in 15cm culture dishes or Cellstack culture chambers (Corning). When cells reached approximately ~75% confluence, CO_2 was dropped to 3% and cells were "triple-transfected" with a 1:1:1 molar ratio of PΔ6, AAVrep.cap, and AAV-ITR containing the transgene or interest using a finely tuned calcium phosphate transfection method at pH~7.03[53]. 24 hours after transfection, media was replaced. 72 hours after transfection cells were harvested by scraping or detached with 10mM EDTA pending transfection vessel. Harvested cells were centrifuged, suspended in tris-buffered saline, and subjected to multiple rounds of freeze-thaw cycles at -80°C followed by sonification (Branson 400, probe tip). Cell debris was removed by centrifugation at 5000xg for 5 minutes, 4°C, followed by treatment with Benzonase and Sodium Deoxycholate (10%). Viruses coding for AAV2.2 were purified using a Hi-Trap Heparin column following manufacturers protocol (GE Healthcare, 5mL, HiTrap HP). Viruses coding for alternative capsid serotypes were loaded onto an iodixanol gradient, ultracentrifuged, and extracted carefully from the 40% band. All viruses were concentrated and desalted

with 100kda concentrator (Corning Spin-X UF Concentrators). Titers were measured using RT-PCR against linearized vector plasmids and against commercially purchased viruses (BioRad SsoAdvanced SYBR). All viruses were sterile filtered (0.22µm) and stored in PBS-Mg with 5% glycerol at -80ºC until need for injection.

2.6.2 Rodent cranial surgery and viral injection

Complications associated with viral vector transduction (3 week expression onset), rodent cranial window surgery (cloudy window and inaccessible after placement), and coelenterazine (CTZ) delivery strategies required multiple factors to be augmented that relate to surgical duration and severity, clarity of window, stability of CTZ, and ability of chemicals to cross the blood-brain-barrier (BBB). Co-developed in an undergraduate thesis by Lin[55], it was determined that chronically implanted cranial windows prevented direct access to the brain for coelenterazine delivery along timelines required for bioluminescent experimentation. The best approach developed for robust viral expression, imaging accessibility, and coelenterazine delivery involved an acute burr hole drilled at precise stereotactic locations over the rodent's barrel cortex followed by an acute cranial window placed on the day of experimentation or an acute craniotomy leaving the brain exposed for manipulation.

Acute surgical burr holes at stereotactic locations were created using custom hardware designed to mount a dental drill onto a precise three axis micromanipulator (Luigs & Neumann). Briefly: after a rodent is anesthetized and pharmacologically prepared for surgery, a small longitudinal incision along the dorsal surface of the head exposes the skull. Aligning the tip of the dental drill bit to bregma allows lateral

(0.4mm) and posterior (0.3mm) coordinates to guide the burr hole location. Drilling the burr hole involves slowly plunging the drill bit tip into the skull while cooling with saline. Successful breach of the skull is verified by a slow leak of CSF over the drill site. Following the burr hole, the drill is swapped for a precise nanoliter injector (Nanoject III, Drummond Scientific) loaded with a microcapillary tube carrying the viral payload. Slow and methodical injections at varied depths prevent viral backflow and minimize damage to the brain surface. After injection, sutures close the incision site and the mouse can recover for 2-3 weeks pending experimental demand. On the day of the experiment, a full cranial window surgery is performed. Before sealing the window, the mouse was injected with 500-1000nL of water-soluble coelenterazine-h [(NanoLight Cat. 3031) dissolved in ringer's solution (do not use PBS) at 10-15mM] along the same stereotactic coordinates that the burr whole was placed. The brain is then sealed with a cranial window or left exposed for further manipulation and prepared for imaging or bioluminescence detection.

2.6.3 Imaging and luminescent detection

All imaging and bioluminescent detection was performed on custom built laser scanning microscopes incorporating at least one of the following commercial femtosecond lasers sources (Ti:Sapphire: Coherent, Chameleon Vision and Mira 720-900nm; Yb:fiber: Amplitude Systems, Satsuma 1030nm). All laser beams were scanned by galvanometric scanners through the back aperture of a range of imaging objectives: Olympus 25x, 1.05NA for high resolution and improved collection efficiency, Olympus 4x, .22NA for centering the animal under the imaging set-up. For hyperspectral

imaging, all three lasers were spatially overlapped, controlled with half-wave plates, and monitored for pulse width and laser power using spectrometers and photodiodes. For bioluminescence detection Coherent Mira at 840nm was used to align the mouse under the objective, and the laser was immediately blocked.

Upon laser excitation (fluorescence) or external stimulation with the laser blocked (bioluminescence), returning light is sent through the objective first crossing a short-pass filter designed to block reflected laser excitation light (but allow most of the visible spectrum to return). The light is then split into four main detection channels with long-pass dichroic mirrors. For hyperspectral, the dichroic mirrors had cut-offs at 495nm, 552nm, and 624nm, which pass light onto a series of angle-tuned bandpass filters (allowing 4 rotations each, for 16 total channels of information per laser). For standard luminescence detection, the filters were fixed and had a center bandpass of 489nm, 550nm, or 595nm; with the fourth emission channel removed to allow synchronization of an externally applied stimulator. Light was detected with gallium arsenide phosphide (GaAsP) photomultiplier tubes (Hammamatsu) controlled with a manually tunable gain box and low-pass filtered at 1 MHz or less pending experimental application (Krohn-Hite). Signal and imaging instrumentation were controlled with National Instruments digital acquisition system (NI-DAQ) and acquired with ScanImage (Vidrio) software in Matlab (MathWorks). Custom image acquisition scripts were written for each respective imaging and detection set-up. Information specific to each modality is commonly included with the header of each image file. For fluorescence, images were processed in ImageJ or custom Matlab script for hyperspectral acquisition[34]. For bioluminescence detection, images were collected as a

series of 1024 pixel line scans. Each line was scanned at approximately 1.2 milliseconds per line, with each pixel representing sampled voltage as a function of pixel dwell time and low pass filtering.

2.6.4 Bioluminescent signal processing and analysis

Each pixel acquired as a function of PMT voltage is considered to represent a combination of bioluminescent emission and/or background electronic noise, shot noise, and other potential sources of noise associated with 2PEF microscopy. Images were initially acquired as a four-channel image stack of 1024x1024 pixel image frames, where z represents the number of frames acquired over time (z x 4 x 1024 x 1024). The images were split into their respective color channels (three center bandpass channels 489nm, 550nm, 595nm) or stimulator TTL pulse channel, and reshaped into long 1D time-series arrays for processing using custom scripts in Matlab or Python. Time-domain average of the resulting signals were initially explored for visualization in Matlab. Window lengths for averaging ranged from 16 to 1024 pixels. These time average windows were explored using common functions in Matlab associated with k-means clustering and principal components analysis with subsequent plots generated for the respective analysis and synced to the 1D vector representing the stimulator output TTL pulse.

Frequency analysis and machine learning was conducted on the raw-reshaped signals in python's Ipython notebook using an extensive list of imported libraries and functions including but not limited to PyWavelets, Scikit-learn, Pandas, Matplotlib, and Numpy[56–62]. Signals were either segmented in to 80millisecond epochs as determined

by the peak locations of the stimulators TTL output pulse, or segmented relevant to the region of interested discussed in the main body of text. The PyWavelets library was employed to explore a series of decompositions, with most analysis focusing on results of the ‘haar’ wavelet. Scipy, Numpy, and Pandas libraries were used to generate statistics on wavelet decomposition and results stored in a pandas DataFrame for further processing. Scikit-learn was thoroughly explored and utilized for both unsupervised and supervised algorithm development and implementation. All models were generated based on existing scripts within this environment. Matplotlib was used to generate initial plots for all analytical results. All plots, tables, graphs, and figures were modified in adobe illustrator with care to preserve all original relationships while formatting the text for publication pre-print.

2.7 References

1. Shanechi, M. M., Chemali, J. J., Liberman, M., Solt, K. & Brown, E. N. A Brain-Machine Interface for Control of Medically-Induced Coma. *PLoS Comput. Biol.* **9,** (2013).
2. Augath, M., Trinath, T., Pauls, J., Logothetis, N. K. & Oeltermann, A. Neurophysiological investigation of the basis of the fMRI signal. *Nature* **412,** 150–157 (2002).
3. Hong, G. & Lieber, C. M. Novel electrode technologies for neural recordings. *Nat. Rev. Neurosci.* 1 (2019). doi:10.1038/s41583-019-0140-6
4. Chaudhary, U., Birbaumer, N. & Curado, M. R. Brain-Machine Interface (BMI) in paralysis. *Ann. Phys. Rehabil. Med.* **58,** 9–13 (2015).
5. Brennand, K. J., Simone, A., Tran, N. & Gage, F. H. Modeling psychiatric disorders at the cellular and network levels. *Mol. Psychiatry* **17,** 1239–53 (2012).
6. M.R., W. *et al.* A prefrontal cortex-brainstem neuronal projection that controls response to behavioural challenge. *Nature* **492,** 428–432 (2012).
7. Mott, M. C., Gordon, J. A. & Koroshetz, W. J. The NIH BRAIN Initiative: Advancing neurotechnologies, integrating disciplines. *PLoS Biol.* **16,** 1–5 (2018).
8. Malliaras, G. G., Deisseroth, K., Wang, H., Rivnay, J. & Fenno, L. Next-generation probes, particles, and proteins for neural interfacing. *Sci. Adv.* **3,** e1601649 (2017).
9. Carmena, J. M. *et al.* Physical principles for scalable neural recording. *Front. Comput. Neurosci.* **7,** 1–34 (2013).

10. Ouzounov, D. G. *et al.* In vivo three-photon imaging of activity of GCaMP6-labeled neurons deep in intact mouse brain. *Nat. Methods* **14,** 388–390 (2017).

11. Guo, Q. *et al.* Multi-channel fiber photometry for population neuronal activity recording. *Biomed. Opt. Express* **6,** 3919 (2015).

12. Lee, S. Y. *et al.* Simultaneous fast measurement of circuit dynamics at multiple sites across the mammalian brain. *Nat. Methods* **13,** 325–328 (2016).

13. Sofroniew, N. J., Flickinger, D., King, J. & Svoboda, K. A large field of view two-photon mesoscope with subcellular resolution for in vivo imaging. *Elife* **5,** (2016).

14. Zipfel, W. R., Williams, R. M. & Webb, W. W. Nonlinear magic: Multiphoton microscopy in the biosciences. *Nat. Biotechnol.* **21,** 1369–1377 (2003).

15. Mitani, A. & Komiyama, T. Real-Time Processing of Two-Photon Calcium Imaging Data Including Lateral Motion Artifact Correction. *Front. Neuroinform.* **12,** 1–13 (2018).

16. Sych, Y., Chernysheva, M., Sumanovski, L. & Helmchen, F. High-density multi-fiber photometry for studying large-scale brain circuit dynamics. *bioRxiv* 422857 (2018). doi:10.1101/422857

17. Gao, P. & Ganguli, S. On simplicity and complexity in the brave new world of large-scale neuroscience. *Curr. Opin. Neurobiol.* **32,** 148–155 (2015).

18. Mazor, O. & Laurent, G. Transient dynamics versus fixed points in odor representations by locust antennal lobe projection neurons. *Neuron* **48,** 661–673 (2005).

19. Ledford, H.; Cryanoski, D. International outcry over genome-edited baby clain.

Nature **563,** 607 (2018).

20. Shimomura, O., Johnson, F. H. & Saiga, Y. Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusan,Aequorea. *J. Cell. Comp. Physiol.* **59,** 223–239 (1962).

21. Inagaki, S. *et al.* Genetically encoded bioluminescent voltage indicator for multi-purpose use in wide range of bioimaging. *Sci. Rep.* **7,** 1–11 (2017).

22. Naumann, E. A., Kampff, A. R., Prober, D. A., Schier, A. F. & Engert, F. Monitoring neural activity with bioluminescence during natural behavior. *Nat. Neurosci.* **13,** 513–20 (2010).

23. Drobac, E., Tricoire, L., Chaffotte, A. F., Guiot, E. & Lambolez, B. Calcium imaging in single neurons from brain slices using bioluminescent reporters. *J. Neurosci. Res.* **88,** 695–711 (2010).

24. Agulhon, C. *et al.* Bioluminescent imaging of Ca2+ activity reveals spatiotemporal dynamics in glial networks of dark-adapted mouse retina. *J. Physiol.* **583,** 945–958 (2007).

25. Baubet, V. *et al.* Chimeric green fluorescent protein-aequorin as bioluminescent Ca2+ reporters at the single-cell level. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 7260–5 (2000).

26. Bakayan, A., Domingo, B., Miyawaki, A. & Llopis, J. Imaging Ca2+ activity in mammalian cells and zebrafish with a novel red-emitting aequorin variant. *Pflugers Arch. Eur. J. Physiol.* **467,** 2031–2042 (2015).

27. Bakayan, A., Vaquero, C. F., Picazo, F. & Llopis, J. Red fluorescent protein-aequorin fusions as improved bioluminescent Ca2+ reporters in single cells and

mice. *PLoS One* **6,** (2011).

28. Yang, J. *et al.* Coupling optogenetic stimulation with NanoLuc-based luminescence (BRET) Ca++ sensing. *Nat. Commun.* **7,** (2016).

29. Iwano, S. *et al.* Single-cell bioluminescence imaging of deep tissue in freely moving animals. *Science (80-.).* **359,** 935–939 (2018).

30. Gealageas, R. *et al.* Bioluminescent properties of obelin and aequorin with novel coelenterazine analogues. *Anal. Bioanal. Chem.* **406,** 2695–2707 (2014).

31. Kim, H. *et al.* Smartphone-based low light detection for bioluminescence application. *Sci. Rep.* **7,** 1–11 (2017).

32. Bares, A. J. *NONLINEAR OPTICAL TOOLS FOR MULTICELLULAR IN VIVO IMAGING*. (2017). doi:https://doi.org/10.7298/X41J97W2

33. Bares, A. J.; Schaffer, C. B.; Tilley, S. Hyperspectral multiphoton microscope for biomedical applications. **1,** (2018).

34. Bares, A. J. *et al.* Hyperspectral Imaging in Live Mouse Cortex Using a 48-Channel Multiphoton Microscope. in *Frontiers in Optics 2016* FTh5D.5 (OSA, 2016). doi:10.1364/FIO.2016.FTh5D.5

35. Pender, M. A. *et al.* Multicolor Genetically-Encoded Calcium-Sensitive Bioluminescent Reporters of Neural Activity for Brain-Machine Interfaces. in *Optics in the Life Sciences Congress* BrTu2B.3 (OSA, 2017). doi:10.1364/BRAIN.2017.BrTu2B.3

36. Karoly, P. J. *et al.* Interictal spikes and epileptic seizures: Their relationship and underlying rhythmicity. *Brain* **139,** 1066–1078 (2016).

37. Hewapathirane, D. S., Dunfield, D., Yen, W., Chen, S. & Haas, K. In vivo

imaging of seizure activity in a novel developmental seizure model. *Exp. Neurol.* **211,** 480–488 (2008).

38. Medeiros, D. C. *et al.* Temporal rearrangement of pre-ictal PTZ induced spike discharges by low frequency electrical stimulation to the amygdaloid complex. *Brain Stimul.* **7,** 170–178 (2014).

39. Kleinfeld, D., Ahissar, E. & Diamond, M. E. Active sensation: insights from the rodent vibrissa sensorimotor system. *Curr. Opin. Neurobiol.* **16,** 435–444 (2006).

40. Petersen, C. C. H. The functional organization of the barrel cortex. *Neuron* **56,** 339–355 (2007).

41. Procházka, A., Kukal, J. & Vyšata, O. Wavelet transform use for feature extraction and EEG signal segments classification. *2008 3rd Int. Symp. Commun. Control. Signal Process. ISCCSP 2008* 719–722 (2008). doi:10.1109/ISCCSP.2008.4537317

42. Letelier, J. C. & Weber, P. P. Spike sorting based on discrete wavelet transform coefficients. *J. Neurosci. Methods* **101,** 93–106 (2000).

43. Quian Quiroga, R. & Nadasdy, Z. Communicated by Maneesh Sahani Unsupervised Spike Detection and Sorting with Wavelets and Superparamagnetic Clustering. **1687,** 1661–1687 (2004).

44. Ergen, B. Comparison of Wavelet Types and Thresholding Methods on Wavelet Based Denoising of Heart Sounds. *J. Signal Inf. Process.* **04,** 164–167 (2013).

45. Tigaret, C. M., Tsaneva-Atanasova, K., Collingridge, G. L. & Mellor, J. R. Wavelet transform-based de-noising for two-photon imaging of synaptic Ca2+ transients. *Biophys. J.* **104,** 1006–1017 (2013).

46. G, M. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **11,** 674–693 (1989).

47. Walker, J. *A Primer on Wavelets and Their Scientific Applications, Second Edition*. **29,** (Chapman and Hall/CRC, 2008).

48. Nerurkar, P., Shirke, A., Chandane, M. & Bhirud, S. Empirical Analysis of Data Clustering Algorithms. *Procedia Comput. Sci.* **125,** 770–779 (2018).

49. Peter J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis . *J. Comput. Appl. Math.* **20,** 53–65 (1987).

50. Shalizi, C. Distances between Clustering , Hierarchical Clustering. *Data Min.* 36–350 (2009).

51. Saxena, A. *et al.* A review of clustering techniques and developments. *Neurocomputing* **267,** 664–681 (2017).

52. Kodinariya, T. . & Makwana, P. . Review on determining number of Cluster in K-Means Clustering. *Int. J. Adv. Res. Comput. Sci. Manag. Stud.* **1,** 90–95 (2013).

53. Morgenstern, P. F., Marongiu, R., Musatov, S. A. & Kaplitt, M. G. Adeno-Associated Viral Gene Delivery in Neurodegenerative Disease. in *Neuroimmune Pharmacology* **793,** 443–455 (2011).

54. Zolotukhin, S. *et al.* Recombinant adeno-associated virus purification using novel methods improves infectious titer and yield. *Gene Ther.* **6,** 973–985 (1999).

55. Lin, K. Developing a robust method for engineering light emitting neurons for brain-machine interface technologies and mapping expression profiles for varying serotypes and injection volumes. (2017).

56. Jones E, Oliphant E, Peterson P, et al. SciPy: Open Source Scientific Tools for Python. (2001). Available at: http://www.scipy.org/ . (Accessed: 15th April 2019)

57. Travis E, O. A guide to NumPy. (2006).

58. Perez, F. & Granger, B. E. IPython: A System for Interactive Scientific Computing. *Comput. Sci. Eng.* **9,** 21–29 (2007).

59. Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **9,** 90–95 (2007).

60. McKinney, W. Data Structures for Statistical Computing in Python. *Proc. 9th Python Sci. Conf.* 51–56 (2010).

61. Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, É. D. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* 2825–2830 (2011).

62. Lee G, Gommers R, Wasilewski F, Wohlfahrt K, O'Leary A, Nahrstaedt H, and C. PyWavelets - Wavelet Transforms in Python. (2006).

CHAPTER 3

3: The Discrete Wavelet Transform in Bioluminescent Signal Processing

3.1 The Discrete Wavelet Transform

The discrete wavelet transform (DWT) is a signal processing tool that is frequently employed for data reduction, filtering, and feature extraction in complex signal analysis. The DWT enables the transformation of discretely sampled time-domain signals into a time-correlated representation of energies corresponding to different frequency bands. Retaining time resolution in the frequency domain is important for understanding signals that have discontinuous components. The properties of the DWT lend itself to many applications in biomedical sciences such as image processing or neuronal spike sorting[1–5]. Although parameters for conducting analysis using the DWT merit some degree of knowledge of the underlying signal, many experimental procedures involve empirical testing of multiple wavelets to achieve optimal analytical results[3,6]. When used in combination with modern data analytical techniques the DWT is well suited for feature extraction and visualization of stochastic signal properties[4,7,8]. The DWT is used in this dissertation as a method of exploration, denoising, and feature extraction of bioluminescent signals emerging from the rodent cortex. The goal is to achieve the best joint time and frequency representation of the signal as possible in order to classify signals in a field that has yet to be established.

3.1.1 Signal analysis and the Fourier transform

Although most signals are analog in nature, instruments for analysis are designed to digitize a signal into a range of discretized samples. The rate at which these signals are discretized is known as the sampling rate, often reported as samples per unit time, f_s. In order to preserve integrity of underlying signal features, well designed instruments should sample analog signals at twice the rate of their highest frequency component (the Nyquist sampling rate)[9]. If the underlying signal propagates a repeatable pattern over a specified window of time, the signal can be considered periodic in nature, where the period (T) is the amount of time required to complete one cycle of its pattern. The frequency can then be expressed as 1/T, commonly reported relative to time in hertz (Hz). Thus, if a period is 10 milliseconds, its frequency is 1/T or 100 Hz. Figure 1 shows an example of individual (**Fig. 3.1a-c**) and the sum (**Fig. 3.1d**) of sinusoids at frequencies 5, 30, or 80 Hz plotted in the time-domain over an interval of one second.

Signals often contain a mix of many frequency components but can ultimately be decomposed into a sum of sine and cosine waves. The frequency representation of a signal can be obtained through Fourier analysis. At its core, the Fourier transform Eq. (1) convolves a time domain signal, *x(t)*, with oscillating sinusoids at different frequencies[10]. Here, the exponential function $e^{-2\pi ift}$ represents a kernel rotating in the complex plane where 2π is circumference, f is the frequency, and t is the amount of time that has passed.

$$\hat{x}(f) = \int_{-\infty}^{\infty} x(t)\, e^{-2\pi ift} dt \qquad \text{(Eq.1)}$$

We can employ one method of Fourier analysis, the Fast Fourier Transform

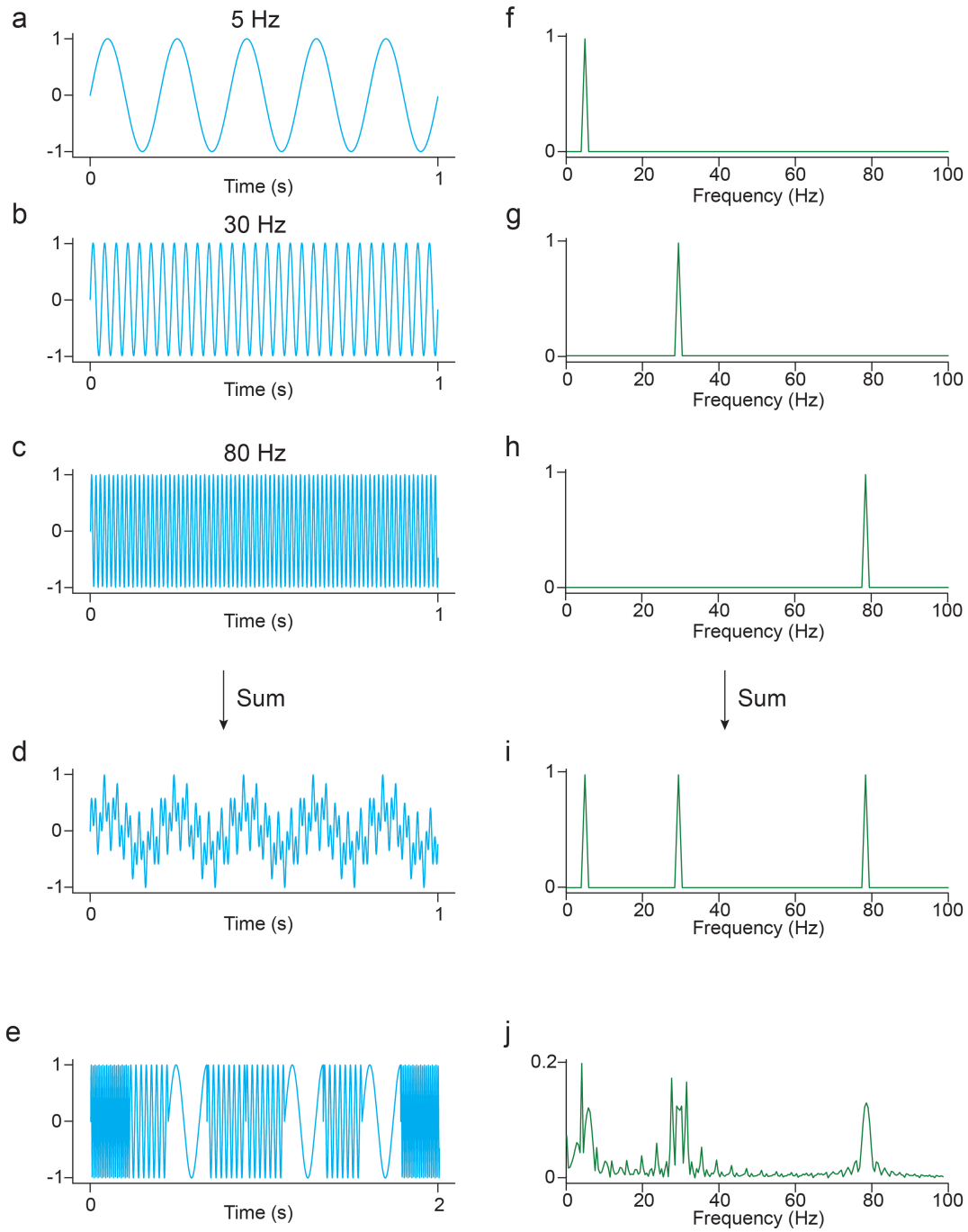


Figure 3.1 Fast Fourier transform of periodic and discontinuous periodic signals
a) 5Hz sine wave, **b**) 30Hz sine wave **c**) 80Hz sine, **d**) and sum of **a-c** all plotted over a 1 second time interval. **f**, **g**, **h, i**) corresponding FFT plotted from 0 to 100Hz for respective signals **a**-**d**. **e**) Discontinuous period appended in semi-random order and used to generate **j**. **j**) FFT of **e**, note that the FFT describes frequency components fairly well, but cannot discern their location along the signal. All plots were produced in Jupyter using Python version 3.7.2 with installed libraries.

(FFT), to capture the presence of each frequency component in a discretely sampled signal by analyzing the output at each level of the frequency convolution (often thought of as a sum of infinitely narrow bandpass filters). The FFT is a fast version of the discrete Fourier transform (DFT) that acts as a linear operator for equally spaced samples, reducing the computation complexity. The FFT of the sinusoids described above are plotted for the first 100Hz in the frequency domain (**Figure 3.1f-i**)[11]. Further equations, theories, and many variations of the Fourier transform are complex in nature and reach beyond the scope of this dissertation. The fundamental equation and demonstrative application of the FFT is presented here to display how time information does not present itself in the frequency domain with traditional Fourier methods. This property is used to motivate the choice of the DWT for analysis of neural luminescent signals.

3.1.2 DWT fundamentals

Since Fourier methods utilize trigonometric functions with infinite time scales for analysis, they cannot be easily employed to resolve locations of abrupt changes (**Fig. 3.1e**) over the entirety of a signal (**Fig. 3.1j**). One could apply the Fourier transform to a small windowed section in the time-domain in order to resolve the location of these changes, but this would sacrifice information about low frequency oscillations that occur outside of the window while effecting other convolution properties[12]. In exchange for time resolution, the DWT sacrifices precise frequency content by utilizing a scalable wavelet as its fundamental unit of analysis. A wavelet is often described as an oscillating signal which has a finite period and zero-mean in the time domain and

therefore can be used as a hybrid between convolution and correlation of multiple signal components[13,14]. For the general case of the continuous wavelet transform (CWT), Eq. (2) has a striking resemblance to the Fourier transform in Eq. (1). Replacing the complex function $e^{-2\pi ift}$ with the wavelet function $\psi_{a,b}(t)$ and outputting a matrix of coefficients $X(a,b)$ as a function of scale a and translation b[15], the wavelet transform can be viewed as a dynamic and tunable kernel for signal analysis.

$$X(a,b) = \int_{-\infty}^{\infty} x(t)\psi_{a,b}(t)dt \qquad \text{(Eq.2)}$$

However, the CWT as defined contains an infinite number of wavelets which can lead to redundancy in data analysis, and possible complications if signal reconstruction is required. Further, wavelets need to be discretized in the time domain in order to use them to analyze discrete signals. Essentially, a discrete wavelet acts as a basis vector of filter coefficients that can be translated along a series of points and scaled to produce a unique output with each translation. The DWT follows the principles of Eq. (2) but employs a scalable discretized wavelet as a filter bank[15]. At each successive level of the DWT's decomposition, the filter is dilated or compressed dependent on the scaling factor a (2,4,8… for the DWT), and translation b[1].

One way to understand the DWT is through visualization of its signal processing framework (**Fig. 3.2**) in contrast to Fourier methods (**Fig. 3.1**). In the presented application, a wavelet acts as a two-component convolution kernel each containing a wavelet-dependent set of discrete filter coefficients related to its shape. In the frequency domain one component serves as a high pass filter, and the other as a low pass filter. In most cases the high pass filter maintains approximately twice the cut-off frequency of

the low pass filter. At each level the low-pass filter output serves as the input for the next level of analysis. The recursive application of a wavelet transform can deconstruct a signal into multiple levels of energy packets that describe an approximate bandpass of frequency components. These components are mapped together into a translation-scale domain relative to the original number of samples (n) and sampling frequency (f_s). Each successive level of deconstruction down-samples the input signal as $f_s/2$, resulting in two time-relative coefficient matrices at half of the original sample number, each retaining a respective split of the frequency content (**Fig. 3.2b**).

To demonstrate this process utilizing the same frequency components from the Fourier example, a discontinuous signal of randomly appended sine waves of 10, 30, and 80 Hz were generated at a 1Khz sampling rate. (**Fig. 3.1e**)[11]. The FFT does well at describing the frequency components that occur in the signal (**Fig. 3.1j**). However, in the frequency domain the FFT has no time resolution beyond the length of the signal itself (we cannot tell when the signals occur, only that they are present during this interval). In contrast, six level decomposition of the discrete wavelet transform utilizing 'db3' from Daubechies family of wavelets illustrates the ability of the DWT to provide both relative time and frequency information (**Fig. 3.2**)[11]. The typical workflow involves: 1. The input of a signal *x(t)*. 2. The convolution of that signal with a wavelet $\psi(t)$ filter bank. 3. The discrete wavelet acts as a two-component filter bank in frequency space, the output includes a high and low pass split of the original signal. 4. The low pass split is used as the input for the next iteration which also scales the wavelet by a factor of two. Note that down sampling by two satisfies the Nyquist sampling rate for these lower frequency components (**Fig 3.2a**)[9].

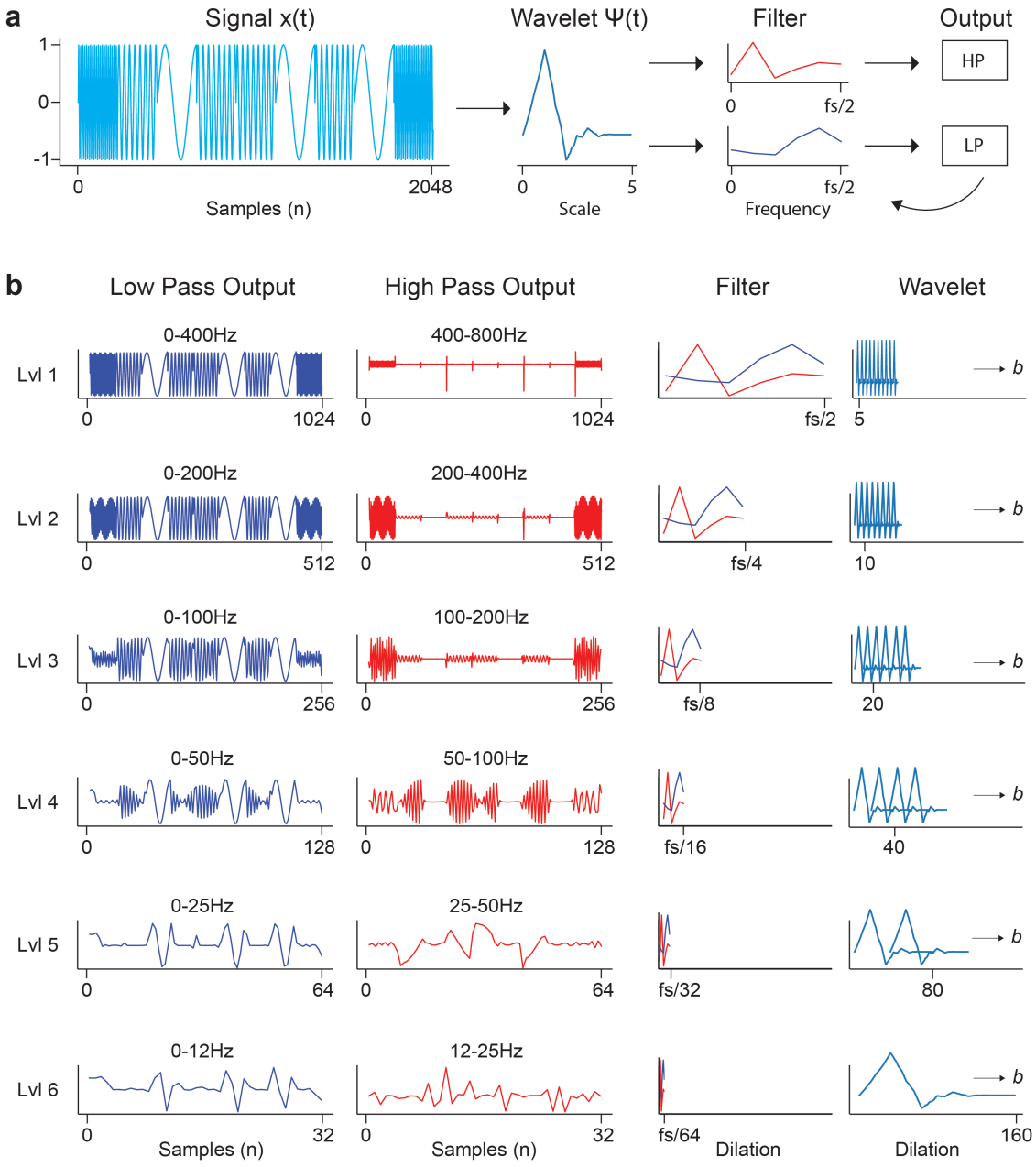


Figure 3.2 Example of the discrete wavelet transform for a discontinuous periodic signal
a) Randomly appended sine waves of 5, 30, and 80Hz as described in (**Fig 3.1a**). The periodic signal x(t) is convolved with a wavelet function $\psi(t)$, the depicted wavelet is Daubechies family of wavelet 'db3'. The discrete wavelet function acts as a dual band filter bank in frequency space. The signal is down sampled as *fs*/2 as the DWT returns the low pass (LP) and high pass (HP) output for further analysis. At each successive level, the LP output is used as an input for the next level. **b**) Full scale demonstration of the DWT on the discontinuous periodic signal in **a** down to six levels of decomposition. Note the emergence of the 80Hz component in HP level 3, the 30Hz component in HP level 4, and the 10Hz component remaining in the lowpass output for level 5-6. The mirrored translation and scale of the frequency filters and mother wavelet are shown alongside the output.

From this wavelet decomposition, one can observe that the wavelet filter bank acts as a time-correlated bandpass filter (**Fig. 3.2b**), allowing the localization of information about the pseudo-frequency content of the signal. Knowing the length of the original signal, sample rate, and wavelet shape one can attain a bandpass window of approximate frequency content. Viewing the different levels of decomposition one can roughly verify the following: level 3 reveals the location of the transitioning 80Hz component in the high pass output; level 4 shows prominence of the 30Hz component in the high pass output; and level 6 reveals elements of the 10Hz component (**Fig. 3.2b**). Alongside the high pass outputs are representative displays of the scaling of the filter components in the frequency domain, and correlated wavelet scale in the time domain. There is a notable tradeoff in resolution with the DWT. As frequency resolution increases, time resolution decreases. At this point, one could infer that frequency content present at different levels of the decomposition may be suitable for generation of signal feature characteristics (entropy, 95th percentile, zero-crossing) or for thresholding certain frequency bands for reconstruction or compression of the signal. The choice of wavelet can also play a role in feature generation by smoothing certain frequency components such as noise[16,17].

In many publications, the choice of the wavelet used for the first level of the transform is generally referred to as the mother wavelet (MWT)[1]. Eq. (3) describes the

$$\psi_{a,b}(t) = \frac{1}{2^{a/2}}\,\psi(2^{a}t - b) \qquad \text{(Eq.3)}$$

general equation for the MWT in the DWT. Here *a* references scale of the wavelet, which usually doubles at each successive level of the DWT, influencing its bandpass

filtering effect in frequency space, and *b* references the translation of the wavelet along the time axis[1,14]. There are many different shapes and scales of wavelets to choose from depending on the underlying features of the signal, the desired frequency resolution, and ultimate level of decomposition[18]. There are few existing principles that can guide the choice of wavelet for signal-type dependent analysis[2]. Generally, time-domain stretched wavelets are better for resolving low frequency features, and compressed wavelets for high frequency features[16]. As underlying signals generated during neural bioluminescence have a complicated mix of frequency components, full exploration of many different types of wavelets may be necessary in many applications but is beyond the scope of detail generated here.

3.2 DWT and Bioluminescent Signal Processing

Following a brief introduction to the DWT, one can see how it might be useful in trying to understand complex signals where the source of the signal is non-stationary and underlying frequency components are generally unknown. Letelier et. al. and Pavlov et. al. demonstrated how the DWT could provide a superior approach over principal components analysis (PCA) by utilizing the wavelet coefficient space for clustering the shapes of neural spike patterns[1,6,19]. Others have explored the ability of wavelets to analyze multi-channel EEG signals through sub-band thresholding and reconstruction or for classification of electrocardiograms[2,3,5]. In the case presented here, bioluminescent signals emerging from the rodent cortex have yet to be classified or clustered into specific waveforms or known frequency patterns such as a the common delta wave in EEG signals[3]. Multichannel neural bioluminescence is quite different from

electrophysiological recordings due to several factors: bioluminescent indicator speed (10-20ms rise time, 500ms decay)[20], detection modality (photon counting PMTs)[21], cellular calcium transient speed (milliseconds following an action potential), depth dependent light scattering (scattering is a function of wavelength and surrounding media), limited spatial resolution (no spatial localization of the signal), and angular incidence of photon flux to the collection head (distant neurons have less impact on signal than close neurons of the same firing behavior). Due to these technical challenges, the properties of the DWT offer a promising approach for signal exploration and possibly setting the stage for a new field of neural recording and classification.

Described in more detail in Chapter 1, 2 and by Pender et al[22], our exploratory dataset is comprised of multicolored bioluminescent signals emerging from virally labeled neurons in an anesthetized rodent's barrel cortex displaying either baseline activity or heightened activity due to a ~10Hz electrical stimulation of the whisker pad. The signals are collected through a high numerical aperture (NA) objective lens and split using dichroic mirrors and bandpass filters (center wavelength at 489nm, 550nm, and 595nm) into three color channels. Each channel detects light with a photomultiplier tube (PMT) which converts the energy of an incident photon into a low-level current output. The output is converted to voltage and amplified, where high frequency noise is commonly low-pass filtered (**Fig. 3.3a**). Signals are represented as a series of appended 1024-pixel lines recorded at 1.2 milliseconds per line. Pixels represent sampled voltage as a function of time; in accordance with the acquisition settings of this signal one pixel is sampled every 117 microseconds; sampling rate for analysis is approximately 850KHz. Trials last on the order of minutes, where multiple stimuli (grey bars) were

applied during the same signal window at differing intervals (**Fig. 3.3b**). Line scans are averaged for presentation in the time domain (**Fig. 3.3b**), they are maintained as raw signals for processing of frequency content (**Fig 3.3c**).

Figure 3.3 displays an overview of bioluminescent detection and signal processing using the DWT for all three-color channels of the first 3 seconds (**Fig. 3.3c**) of the first 10Hz stimulus outlined with the widest grey box in **Figure 3.3b**. Each row is labeled by the corresponding upper band of approximate frequency content for the high pass component, where the low pass band contains frequency components that are one octave below the row label. Although axes and labels have been removed to display more content in one figure, many insights can be gained from this scale-translation domain of the first stimulus triggered neural bioluminescent signals decomposed down to 14 levels with a 'haar' wavelet. First, the 489nm and 550nm color channels appear to contribute greater to emergent signals in both the time and frequency domain. Second, every level of the high pass filter bank appears to contribute to the emergent signal; it is readily apparent where the onset of the first stimulus occurred, which is marked by a short vertical line at the bottom of the figure. Third, the low pass components in the sub-kHz regime start to reveal a feature that mirrors the rise and decay of a typical bioluminescent indicator, suggesting an initial burst of neural activity characteristic of a similar type of stimulus[23,24]. Fourth, the frequency content remains elevated and oscillatory during and after this initial burst of activity. These features suggest that there may be different populations of neurons active throughout this trial, and that frequency components could help to describe that activity.

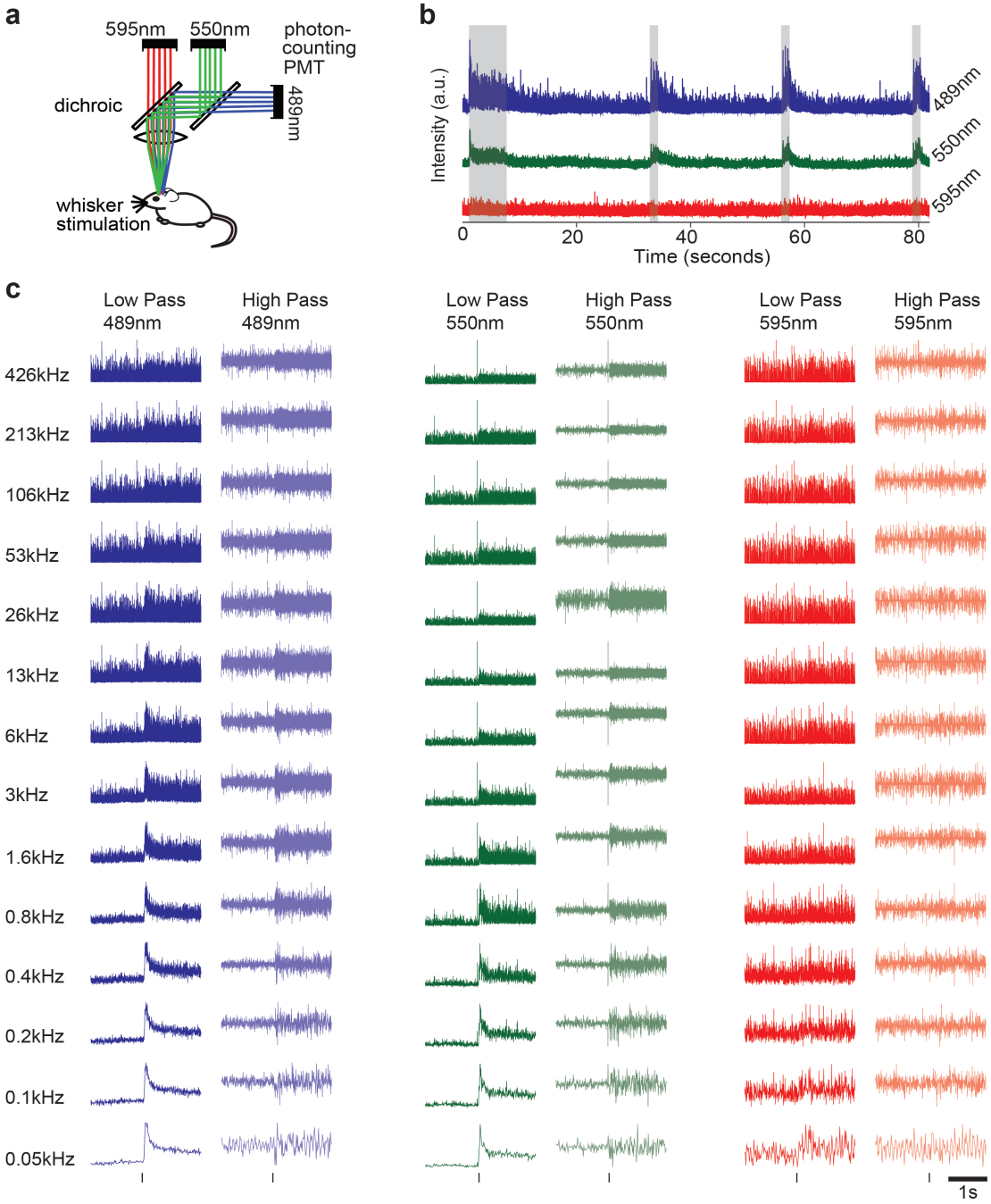


Figure 3.3 The discrete wavelet transform for neural bioluminescent signal analysis
a) Low detail schematic of the optical detection set-up used to collect bioluminescent signals from the rodent cortex. Signals are generated by a reaction inside the neural cells of the mouse, each cell contains combinations of colored luminescent proteins. Light emitted is split first by a dichroic mirror, and then bandpass filter to a specified wavelength (489nm, 550nm, and 595nm note the center wavelength). **b**) Example of a full stimulus trial with multiple 10Hz stimuli market in light grey boxes. The 10Hz stimulus is applied for different durations, the mouse is allowed time in between stimuli to recover. **c**) A full 'haar' wavelet DWT carried out on the first 3 seconds of the first stimulus near time t=0 in **b**. Stimulus onset is marked by a short vertical line at the bottom of the image. Rows are labeled by high pass upper band frequencies; low pass frequencies are reference 1 octave below the row label. x-axis represents approximate time or translation; y-axis represents magnitude in relative frequency/scale domain.

Although not fully presented in this text, various wavelet types and scales were explored in processing different regions of these stimulus evoked neural signals. The general conclusions drawn in **Figure 3.3** are quite common for employment of most discrete wavelet families and windowed regions. While drawing a direct conclusion about specific neurons contributing to this signal is likely impossible, it is probably also unnecessary. Since we are collecting photons of characteristically slow luminescent probes from many neurons of different firing rates, and from many incidence angles on the objective, a reductionist approach will likely experience more experimental error than conclusive results. We are sampling light, which travels fast, so we should expect high information density and consider an approach that merits large scale data processing.

3.2.1 DWT for denoising neural luminescent signals

Although it was not discussed in the introductory section, the DWT has interesting properties in that its coefficients can be thresholded in a variety of ways in order to denoise or compress the data while preserving features. So long as the coefficient matrix remains the same length, reconstruction is the inverse of the DWT, a characteristic of this technique that allows sharp details to be preserved while possibly removing noise at every level of decomposition. Although quite standard in most imaging experiments, noise in this particular application needs to be carefully considered as we are aiming to explain neural activity from extremely dim signals. Common sources of noise include white noise, pink noise, and electronic noise, amongst others[25]. Most significantly, we need to explore the signals prior to the

stimulus to better understand whether the mouse may be emitting baseline luminescence. In order to understand the effect of noise that we could not control for with digital filtering, we used signal collected in the absence of a rodent as a baseline for thresholding and analysis during wavelet decomposition. We employed the general method describe by Quiroga et al to threshold wavelet frequency bands of the baseline signal at all levels (**Fig. 3.4**)[26].

$$4 * \sigma = 4 * MAD = 4 * median\left\{\frac{|x|}{0.6745}\right\} \qquad \text{(Eq.4)}$$

From Eq. (4) we used the median absolute deviation (MAD) thresholding method where x is the absolute median of the signal in a specific frequency sub-band. To ensure we maximized the capture of baseline signal, the measurements from multiple baseline readings were averaged. Although only the MAD results are reported in **Figure 3.4**, we explored many other options including Stein's unbiased risk estimation *Rigresure* and the universal threshold *Sqtwolog*[27]. This exploration was an important exercise, as it became apparent that the mouse was possibly displaying stochastic neuronal firing consistent with baseline activity measurements in an anesthetized mouse at different frequency bands[28].

Figure 3.4 illustrates the results of baseline thresholding near the site of the expected luminescent signals due to whisker pad stimulation. It is apparent that a significant amount of signal exists well above baseline. Further, the activity deviations in the high frequency bands implied the importance of maintaining these features for accurate signal analysis. Although the data suggests that overall variance is fairly low in higher frequency regions; thresholding and analysis suggests these regions need to be

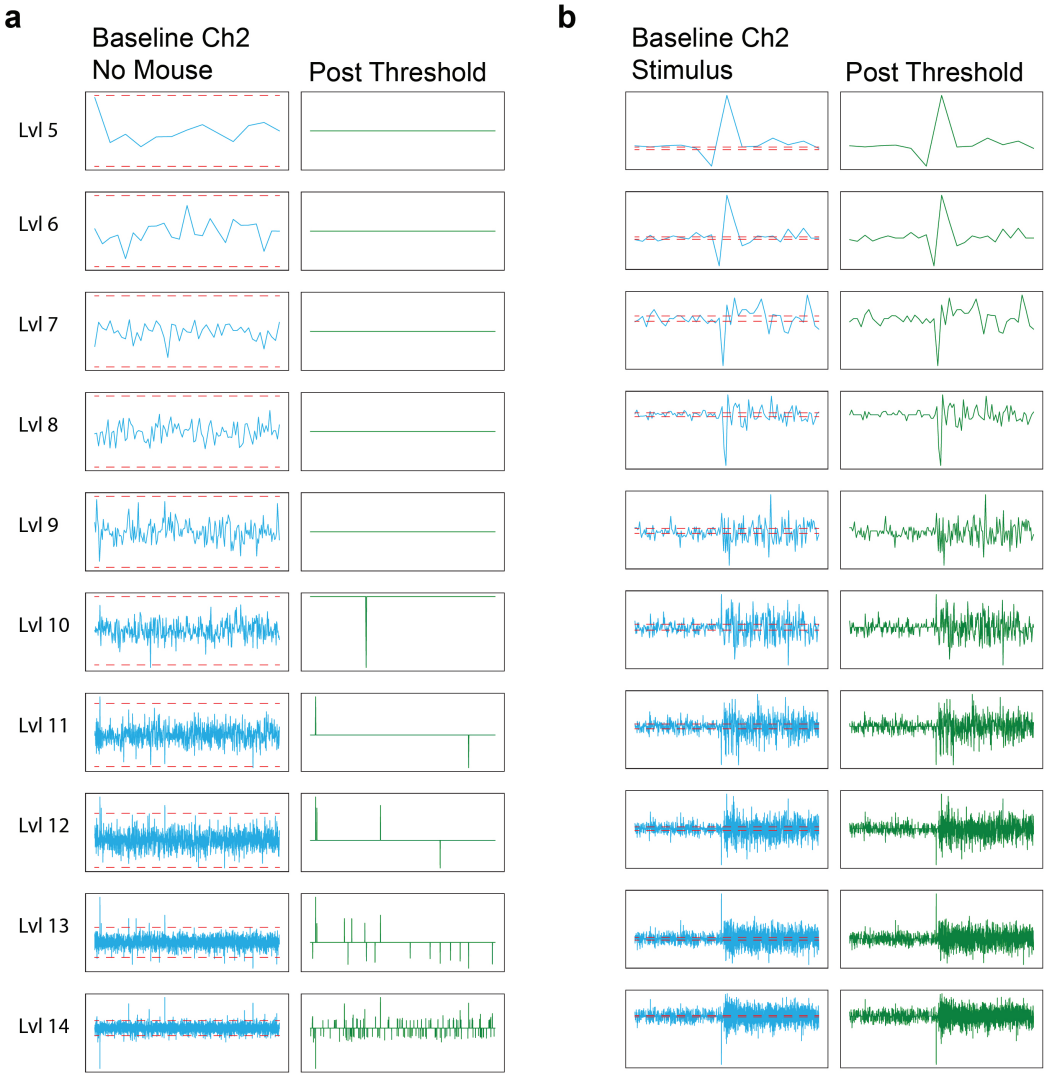


Figure 3.4 Thresholding coefficients with the discrete wavelet transform
a) Level 5-14 DWT decomposition displaying the original detail coefficients for baseline noise on the left "Baseline Ch2 No Mouse." 4*MAD (median absolute deviation) was used to set thresholds unique to each level of decomposition (dashed red lines). Column 2 "Post Threshold" displays detailed coefficients after applying the hard threshold describe in column 1. **b**) Threshold values establish in **a** were used to threshold the same the 550nm channel from Figure 3.3. "Post threshold" values suggest that baseline bioluminescence activity may exist in the anesthetized mouse prior to stimulus.

kept for further exploration. Therefore, we decided to use the DWT as a tool for gathering features from raw data. Rather than averaging or windowing in the time domain to generate features, the raw signal was segmented based on applied stimulus epochs.

3.2.2 DWT for feature generation

After thoroughly exploring the frequency space via plotting, denoising, and examining statistics, it was determined that all frequency bands may be useful in determining characteristics of the underlying signals. Since we have limited information about the stimulus applied, we may be able to use this frequency content to generate a classifier. To generate features for the classifier, we parsed into 80 millisecond epochs (approximately the width of one stimulus cycle), keeping data at all levels, extracting features while simultaneously reducing data dimensionality. Features such as entropy, variance, mean, 0-crossing, and percentile scores were extracted from each frequency sub-band of each epoch, generating a new feature vector of varying length depending on which wavelet was used. We then used this data to test different unsupervised learning methods (i.e. hierarchical clustering, k-means clustering, and principal components analysis) and classification with supervised algorithms. Consult Chapter 2 for further detail.

3.3 References

1. Letelier, J. C. & Weber, P. P. Spike sorting based on discrete wavelet transform coefficients. *J. Neurosci. Methods* **101,** 93–106 (2000).
2. Zhang, Z., Telesford, Q. K., Giusti, C., Lim, K. O. & Bassett, D. S. Choosing wavelet methods, filters, and lengths for functional brain network construction. *PLoS One* **11,** 1–24 (2016).
3. Al-Qazzaz, N. K., Mohd Ali, S. H. Bin, Ahmad, S. A., Islam, M. S. & Escudero, J. Selection of mother wavelet functions for multi-channel EEG signal analysis during a working memory task. *Sensors (Switzerland)* **15,** 29015–29035 (2015).
4. Procházka, A., Kukal, J. & Vyšata, O. Wavelet transform use for feature extraction and EEG signal segments classification. *2008 3rd Int. Symp. Commun. Control. Signal Process. ISCCSP 2008* 719–722 (2008). doi:10.1109/ISCCSP.2008.4537317
5. De Chazal, P., Celler, B. G. & Reilly, R. B. Using wavelet coefficients for the classification of the electrocardiogram. *Annu. Int. Conf. IEEE Eng. Med. Biol. - Proc.* **1,** 64–67 (2000).
6. Pavlov, A., Makarov, V. A., Makarova, I. & Panetsos, F. Sorting of neural spikes: When wavelet based methods outperform principal component analysis. *Nat. Comput.* **6,** 269–281 (2007).
7. Ghazali, K. H., Mansor, M. F., Mustafa, M. M. & Hussain, A. Feature extraction technique using discrete wavelet transform for image classification. *2007 5th Student Conf. Res. Dev. SCORED* 5–8 (2007).

doi:10.1109/SCORED.2007.4451366

8. Shinomoto, S. *et al.* Relating neuronal firing patterns to functional differentiation of cerebral cortex. *PLoS Comput. Biol.* **5,** (2009).

9. Landau, H. J. Sampling, Data Transmission, and the Nyquist Rate. *Proc. IEEE* **55,** 1701–1706 (1967).

10. Oberst, U. The Fast Fourier Transform II. **46,** 2005 (2005).

11. Lee G, Gommers R, Wasilewski F, Wohlfahrt K, O'Leary A, Nahrstaedt H, and C. PyWavelets - Wavelet Transforms in Python. (2006).

12. Zhu, X., Beauregard, G. T. & Wyse, L. L. Real-time signal estimation from modified short-time fourier transform magnitude spectra. *IEEE Trans. Audio, Speech Lang. Process.* **15,** 1645–1653 (2007).

13. Polikar, R. The Engineer's Ultimate Guide to Wavelet Analysis. *Wavelet Tutor.* (1999). doi:2200/20100707002841225T

14. Fedi, M., Primiceri, R., Quarta, T. & Villani, A. V. Joint application of continuous and discrete wavelet transform on gravity data to identify shallow and deep sources. *Geophys. J. Int.* **156,** 7–21 (2004).

15. G, M. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **11,** 674–693 (1989).

16. Graps, A. An introduction to wavelets. *Choice Rev. Online* **30,** 30-0976-30-0976 (2013).

17. Tigaret, C. M., Tsaneva-Atanasova, K., Collingridge, G. L. & Mellor, J. R. Wavelet transform-based de-noising for two-photon imaging of synaptic Ca2+ transients. *Biophys. J.* **104,** 1006–1017 (2013).

18. Wasilewski, F. Wavelet Browser. Available at: http://wavelets.pybytes.com/. (Accessed: 4th April 2019)

19. Hulata, E., Segev, R. & Ben-jacob, E. A method for spike sorting and detection based on wavelet packets and Shannon's mutual information. *J. Neurosci. Methods* **117,** 1–12 (2002).

20. Shimomura, O., Johnson, F. H. & Saiga, Y. Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusan,Aequorea. *J. Cell. Comp. Physiol.* **59,** 223–239 (1962).

21. K, H. P. K. Principle of Photon Counting - Chapter 6. *Hamamatsu Photonics K. K.* (2007).

22. Pender, M. A. *et al.* Multicolor Genetically-Encoded Calcium-Sensitive Bioluminescent Reporters of Neural Activity for Brain-Machine Interfaces. in *Optics in the Life Sciences Congress* BrTu2B.3 (OSA, 2017). doi:10.1364/BRAIN.2017.BrTu2B.3

23. Petersen, C. C. H. The functional organization of the barrel cortex. *Neuron* **56,** 339–355 (2007).

24. Kleinfeld, D., Ahissar, E. & Diamond, M. E. Active sensation: insights from the rodent vibrissa sensorimotor system. *Curr. Opin. Neurobiol.* **16,** 435–444 (2006).

25. Jonas, M. Optical detectors, noise, and the limit of detection. Available at: http://measurebiology.org/wiki/Optical_detectors,_noise,_and_the_limit_of_detection. (Accessed: 4th April 2019)

26. Quian Quiroga, R. & Nadasdy, Z. Communicated by Maneesh Sahani Unsupervised Spike Detection and Sorting with Wavelets and Superparamagnetic

Clustering. **1687,** 1661–1687 (2004).

27. Ergen, B. Comparison of Wavelet Types and Thresholding Methods on Wavelet Based Denoising of Heart Sounds. *J. Signal Inf. Process.* **04,** 164–167 (2013).

28. Hildebrandt, K. J., Sahani, M. & Linden, J. F. The Impact of Anesthetic State on Spike-Sorting Success in the Cortex: A Comparison of Ketamine and Urethane Anesthesia. *Front. Neural Circuits* **11,** 1–12 (2017).

CHAPTER 4

4: Bioluminescent Markers of Neural Activity

4.1 Abstract

Fusion of photoprotein aequorin (apoaequorin) to a fluorescent protein partner is one method to alter bioluminescent characteristics of coelenterazine substrate catalysis. Here, we generate novel bioluminescent protein fusion constructs with the goal of mapping neural activity through genetic spectral barcoding; where indicator speed and color are salient. Kinetically altered fusion constructs are created by exchanging apoaequorin with alternative enzymatic photoprotein donors from *A. Macrodactyla* (apo-Amac) and *O. Longissima* (apo-obelin). Bright and spectrally distinct fusion constructs are created by swapping protein fusion partners from existing fluorescent species. Where fusion partners are energetically unfavorable, protein subunits are truncated or new tandem dimer fluorescent proteins are created. Blue-shifted fluorescent protein fusions have enhanced relative luminescent output up to two-fold greater for mTFP1-aequorin compared to existing eGFP-aequorin. Photoprotein donor Obelin from *O. Longissima* enhances the kinetic profile compared to existing fusion constructs based on aequorin. We also report the furthest red-shifted bioluminescent indicators based on modified mScarlet and mRuby3 fluorescent proteins. The compiled result is a palette of spectrally separable, bright, and fast bioluminescent proteins that could enable enhanced resolution when applied to chromatic luminescence as a cellular level readout system of neural activity. Refined characterization of the developed indicators may also help elucidate complex mechanisms of bioluminescence in fusion proteins.

4.2 Introduction

It is well known that regulation of cellular Ca^{2+} is associated with many homeostatic mechanisms of in vivo physiology[1–3]. In neurons, Ca^{2+} is well documented as a secondary messenger of neural activity, and is involved in nearly every reaction that occurs within the brain[1,4,5]. In neuronal signaling, Ca^{2+} influx directly follows an action potential with temporal oscillations ranging from 1Hz to 1Khz but can be even faster in certain microdomains. The calcium fluctuations associated with action potentials can range from low nanomolar to 10-100uM or more dependent on the localization of the signal and distance of propogation[4,6].

Bioluminescent photoproteins such as apo-aequorin from *A. Victoria* and apo-obelin from *O. Longissima* can emit light in response to Ca^{2+} changes when pre-charged with its "luciferin" chromophore coelenterazine (CTZ)[7,8]. In the charged state, these molecules are commonly referenced as aequorin (AEQ) or obelin (OBE), and have calcium binding affinities in the low nanomolar range[9]. Previously, it has been shown that fusion proteins based on AEQ or OBE could report Ca^{2+} signals from stimulated neurons in culture[10]. Many strategies have been explored to tune the calcium-dependent spectral emission profiles of OBE and AEQ. On one end, mutating residues associated with the CTZ binding site can shift the spectral profile of AEQ or OBE[9,11,12]. On the other end, emission characteristics can be augmented through interchange of the CTZ substrate; where each substrate possesses a range of photoprotein-dependent kinetic and chromatic properties[13,14]. Since the photoprotein itself is only responsible for the enzymatic degradation of CTZ to produce light, nomenclature can be rather perplexing given the combinatorics of mutated OBE and AEQ variants coupled to the range of

CTZ substrates available.

Adding further complication, many fluorescent proteins (FPs) have been fused to AEQ variants to mimic the efficiency of resonant energy transfer noted in jellyfish; where many fusion based mechanisms have yet to be fully elucidated[15–18]. In one example, utilizing identical CTZ substrates in solution, AEQ (470nm emission) and OBE (480nm emission) can swap emission profiles by mutating one residue within their active site[11]. However, when fused to green fluorescent protein (GFP), GPF-AEQ has an enhanced red-shifted output and transfer mechanism when compared to GPF-OBE[19]. These features suggest that AEQ's interaction with GFP has enhanced resonance energy transfer (RET) by a calcium-dependent interaction which is more pronounced than RET between OBE and GFP[17,20]. In a study by Alakhov et al., this interaction was responsible for time-dependent spectral changes during the bioluminescent emission process[19]. Here, we keep this concept in mind while exploring multiple avenues that may be useful for indicator design in calcium-dependent neural bioluminescence applications.

To utilize bioluminescence in chromatic neural interfaces where time-dependent spectral components are salient, many behaviors of the bioluminescent protein fusion constructs need to be considered in more detail and in a broader scope. With the goal of increasing information gain using a non-spatially resolved imaging modality, bioluminescent indicators must have spectrally separable and time-invariant emission profiles. In order to resolve multiple events from fast-firing neurons, indicators need enhanced speed and sensitivity to report fast calcium transients. Here, we build off of existing indicators designed by Bakayan et al through testing new fluorescent protein

fusion partners and donor species[21,22]. Further, we characterize new and existing indicators with the luciferin substrate that we found to perform optimally in vivo: water-soluble CTZ-H (NanoLight-Prolume)[14].

4.3 Results

4.3.1 Enhanced relative luminescence output from blue-shifted aequorin fusions

The fusion of aequorin to a fluorescent protein enhances the light emitting properties of the fusion construct compared to aequorin alone. Although a range of alternative mechanisms have been suggested, reports focus primarily on protein stability and RET [22–24]. The principles of RET between two luminescent species are most commonly described by the Förster equation, and depend on many factors such as distance, orientation, and spectral overlap between two molecules[25,26]. Here, the distance between two molecules in solution is forced into proximity through a flexible amino acid linker[22]. To enhance the relative luminescence output (RLO) across the blue end of the visible spectrum, we fused AEQ and OBE to a range of blue absorbing fluorescent proteins: mTagBFP2, mTFP1, mCerulean, and eCFP where spectral overlap may be suitable for energy transfer (**Supplemental Fig. 4.4a**). The bioluminescent emission profiles of these newly created fusion constructs charged with CTZ-H were collected on an Ocean Optics QE65000 spectrometer in response to saturating calcium (**Fig. 4.1a**), where they retain similar emission profiles to the native fluorescent protein. We also tested the RLO of each protein normalized to its concentration (as determined by peak fluorophore absorbance), reported as a relative percentage of eGFP-AEQ's bioluminescence output (**Fig. 4.1b**). Results suggest that blue-shifted fusion constructs

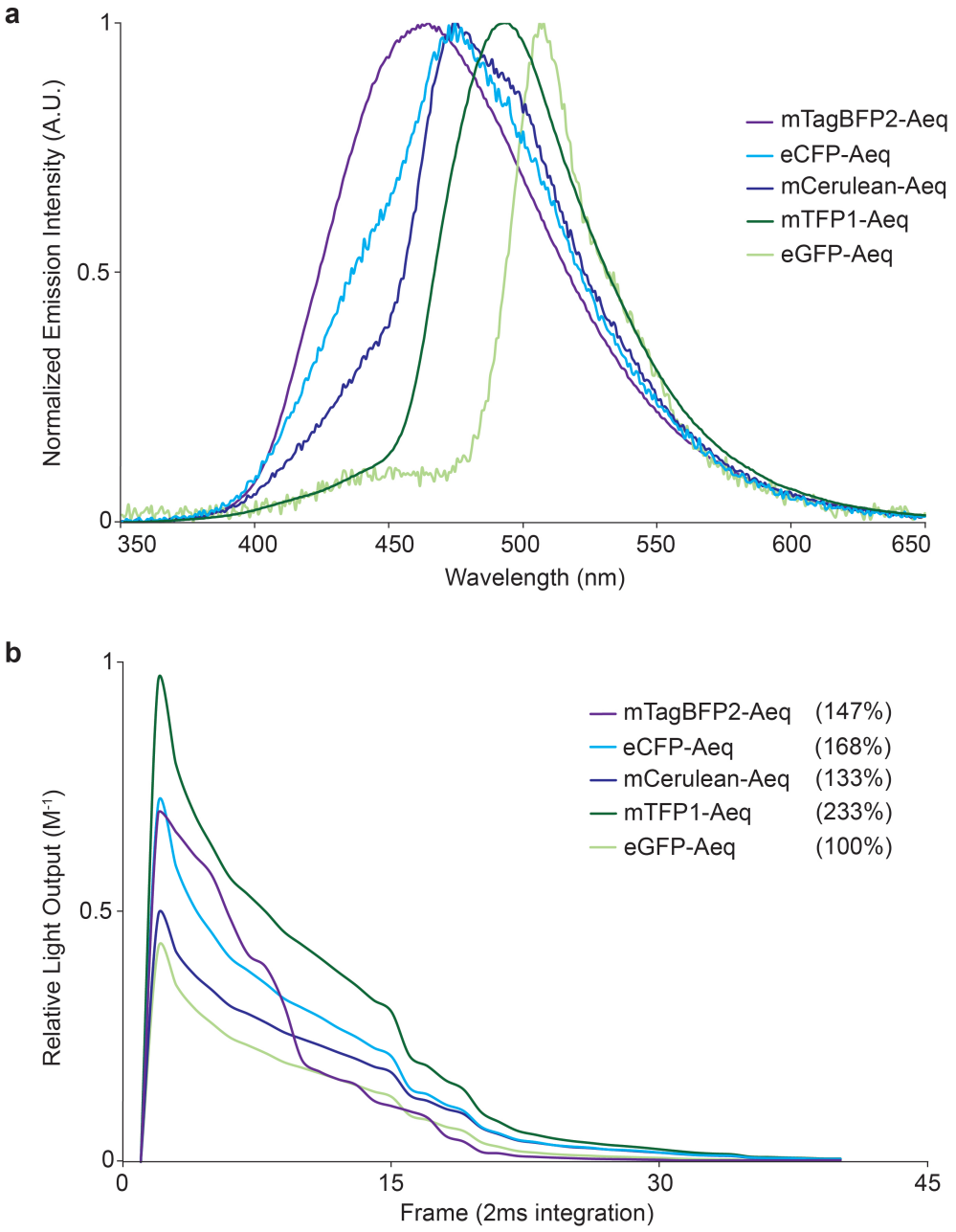


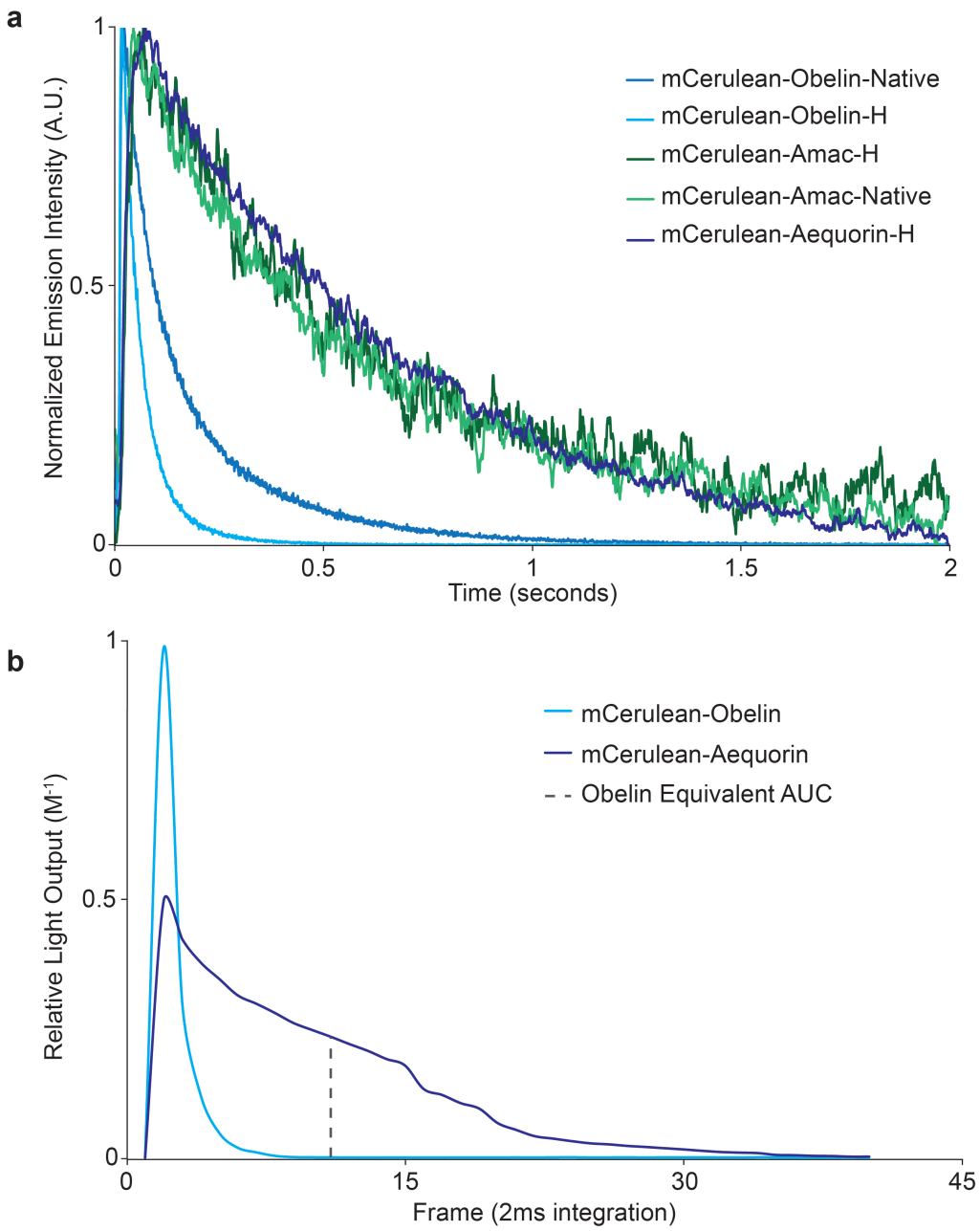
Figure 4.1 Blue-shifted bioluminescent spectra and relative light output
a) Normalized bioluminescent emission spectra for all novel blue-shifted protein fusions plotted with original eGFP-Aeq. Spectra sampled in response to saturating Ca^{2+} (200mM) at room temperature with 100ms integration (Ocean Optics QE65000) after overnight incubation with CTZ-H. **b**) Relative light output normalized to protein concentration, light output is calculated as mean gray value per frame collected with Basler AG CMOS camera; average of three trials. Total light collected over entire time course is stated as a (%) relative to eGFP-Aeq.

create brighter indicators that have greater light output on a "per fluorescent molecule" basis, where mTFP1-AEQ has approximately twice the RLO of eGFP-AEQ.

4.3.2 Improved kinetic response from Obelin fusion constructs

Obelin from *O. Longissima* has been reported to produce a much faster Ca^{2+} dependent bioluminescent response profile than that of AEQ[9,13]. Further, there have been very few studies which shift the spectral profile of OBE through a fluorescent protein fusion partner or elucidate associated mechanisms involved in RET[16,19]. As mentioned earlier, OBE does not have a calcium enhanced RET transfer to GFP in solution, making it an ideal candidate for developing sensors with time-invariant spectral emission profiles. Here, a variety of protein acceptors were fused to OBE, as well as another species of aequorin - *A. Macrodactyla (AMAC)* – attempting to enhance the spectral diversity and kinetic response profile over existing bioluminescent constructs. Testing the kinetic response profile of new fusions proteins, we utilized an SFM-400 stopped-flow rapid mixer, collecting bioluminescent emission in response saturating calcium. **Figure 4.2a** displays the stopped-flow acquired emission spectra of new constructs based on mCerulean fused to OBE, AEQ, and AMAC utilizing the standard fusion linker as described by Bakayan et al[22]. We can see that mCerulean-OBE charged with CTZ-H produced the fastest response profile with a $K_{1/2rise}$ of ~14ms and $K_{1/2decay}$ of ~66ms; comparatively faster than AMAC and AEQ fusions. The improvement of both rise and decay rates could be useful in separating signals from neurons that have fast inter-spike intervals.

Figure 4.2b shows the RLO of mCerulean-OBE in comparison to mCerulean-



	OBE-Native	OBE-H	Amac-H	Amac-Native	Aeq-H
$K_{1/2rise}$ (ms)	14	14	25	26	18
$K_{1/2decay}$ (ms)	110	66	379	418	398

Figure 4.2 Kinetically altered fusion proteins with Obelin or A Macrodactyla
a) Normalized emission spectra collected in SFM 400 stopped-flow mixer; aequorin spectra underwent exponential smoothing to improve plot quality (alpha=0.9). Kinetic improvements are substantial for Obelin with different CTZ types, not as great for Amac. **b**) Comparison of Obelin vs Aequorin mCerulean fusion integrated luminescent intensity over time.

AEQ, measured as previously described in **Figure 4.1b**. Here, a dashed line represents the integrated intensity of mCerulean-OBE relative to an equivalent integrated intensity of mCerulean-AEQ. Although the acquisition system used to collect RLO is based on a less sensitive photodetector than stopped-flow mixing (and only able to capture the peak output), the importance of speed and brightness is well illustrated. Most notably, mCerulean-OBE concentrates a majority of its RLO over a very short time frame; an ideal feature for separating calcium spikes following neural action potentials.

4.3.3 Red-shifted indicators based on obelin and aequorin

New fluorescent proteins are developed every year, and they often display enhanced photophysical characteristics that improve RET acceptor-based properties such as enhanced brightness, molar extinction coefficient, or donor-acceptor spectral overlap (**Supplemental Fig 4.4b**). These properties lend themselves to being tested as fusion protein partners of AEQ and OBE. Here we fused recently developed mScarlet and mRuby3 to AEQ and OBE utilizing the standard linker outlined by Bakayan et al[22,27,28]. Noting that red-shifting was rather limited (**Table 1**), we created tandem monomers out of mScarlet and mRuby3 by fusing two copies via a linker used in the development of tdimer (GHGTGSTGSGSS)[29]. The tandem monomer strategy provided a roughly 2-fold improvement in transfer ratio as calculated by peak-to-peak acceptor-to-donor bioluminescence emission ratio. To improve transfer, we truncated the linker between AEQ and mRuby3, creating the furthest red-shifted aequorin based indicator to date. We also created a fast red-shifted indicator by fusing Obelin to tdTomato (**Table 1**).

These indicators also appear to demonstrate other unique spectral properties. In

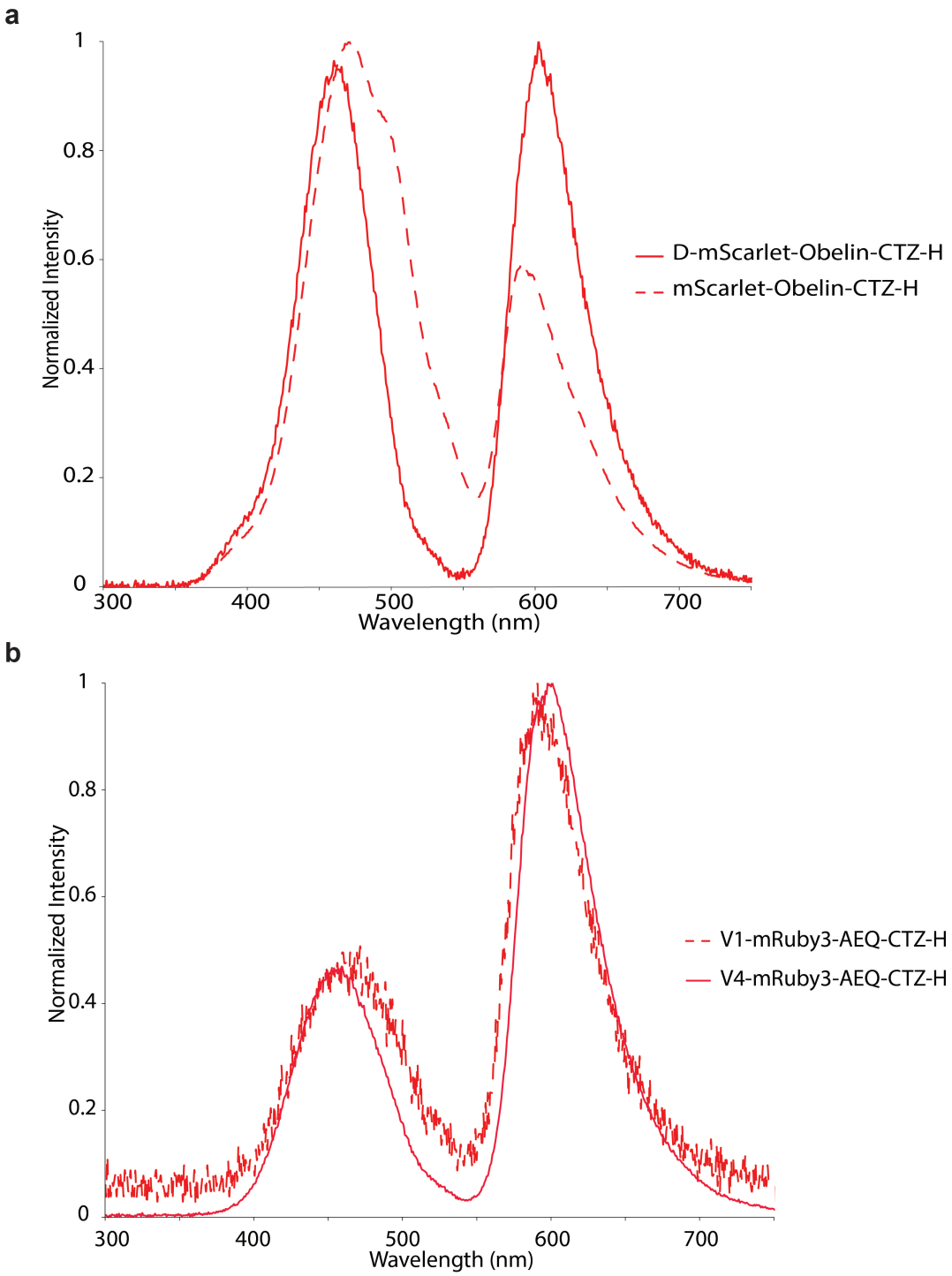


Figure 4.3 Red-shifted indicators based on mRuby3 and mScarlet
Normalized bioluminescent emission spectra of novel red-shifted protein fusions from Table 1. Spectra were sampled in response to saturating Ca^{2+} (200mM) at room temperature with 100ms integration (Ocean Optics QE65000) after overnight incubation with CTZ-H. **a**) Profile for dimer and monomer made from mScarlet and OBE. **b**) Profile for truncated monomers made from mRuby3 and AEQ.

the case of mScarlet, fusing the fluorescent protein to OBE caused a red-shift in the donor peak (472nm) and blue-shift in the acceptor peak (588nm) compared to the otherwise identical mScarlet-AEQ construct (599nm and 457nm respectively). A similar shift was notable in the peak emission of the donor for the dimer constructs based on mScarlet (447nm for AEQ and 459nm for OBE), but the acceptor remained relatively similar near 600nm. These properties merit further experimentation to determine RET enhancement by calcium dependent mechanisms.

Table 1: Red-shifted bioluminescent fusion constructs

Construct Name	D Emission (λ nm)	A Emission (λ nm)	Peak A/D
Dimer-mRuby3-AEQ-CTZ-H	**456.4** +/- 7.3	**592.3** +/- 5.9	0.86 +/- 0.06
Dimer-mScarlet-AEQ-CTZ-H	**447.3** +/- 6.3	**603.8** +/- 5.0	0.91 +/- 0.05
Dimer-mScarlet-Obelin-CTZ-H	**459.0** +/- 16.1	**599.0** +/- 9.5	0.93 +/- 0.07
mScarlet-AEQ-CTZ-H	**457.1** +/- 3.0	**598.9** +/- 1.2	0.51 +/- 0.04
mScarlet-AEQ-CTZ-Native	**460.2** +/- 1.9	**599.0** +/- 1.0	0.51 +/- 0.02
mScarlet-Obelin-CTZ-H	**472.2** +/- 3.0	**588.4** +/- 1.1	0.49 +/- 0.02
mScarlet-Obelin-CTZ-Native	**472.1** +/- 1.4	**588.1** +/- 1.8	0.47 +/- 0.02
tdTomato-Obelin-CTZ-H	**471.4** +/- 4.8	**582.0** +/- 1.6	1.84 +/- 0.03
V1-mRuby3-AEQ-CTZ-H	**455.1** +/- 8.2	**593.7** +/- 7.8	2.11 +/- 0.23
V2-mRuby3-AEQ-CTZ-H	**451.4** +/-2.5	**602.6** +/- 3.0	2.29 +/- 0.11
V3-mRuby3-AEQ-CTZ-H	**451.4** +/- 9.4	**599.2** +/- 5.4	2.44 +/- 0.17
V4-mRuby3-AEQ-CTZ-H	**454.8** +/- 2.5	**600.0** +/- 1.4	2.24 +/- 0.06

4.4 Discussion

In this study we have confirmed the ability of different blue and red fluorescent proteins to serve as RET partners of OBE and AEQ. Indicators presented here seem to follow the general guidelines of RET as outlined by the Förster equation, where donor acceptor overlap, distance, and perceived extinction coefficient (not measured) likely influence the transfer ratio. We've also demonstrated that OBE can act as a fast fusion partner across multiple FP acceptor species. Its ability to report calcium with a high RLO over a

very short period is desirable for future use as a neural activity reporter. Further, as bioluminescent photoproteins already emit in the blue wavelength regime, limited studies exist on blue-shifting or "blue-stabilizing" photoprotein indicators; there has been little need until the development of applications such as the BMI described in Chapter 2. Although blue light is troubled by scattering in biological tissue, these stabilized blue shifted indicators will be extremely useful for reporting robust activity from the surface of the cortex or at depth utilizing an implantable fiber.

4.5 Conclusions and Future Directions

Due to its speed, red-shifting ability, and concentration of RLO over a short time frame, Obelin could serve as a useful reporter of neural activity and has the promise of better resolving neural action potentials in non-spatially resolved imaging modalities. We have also demonstrated a diverse palette of spectrally unique indicators which could be used with different genetic targeting strategies in chromatic bioluminescence applications. A combination of both or many other kinetic variants could also be used to add information density to a signal, as slow and fast components could report calcium microdomain changes that may be significant in understanding neural activity. Although not explored here, extended studies are recommended to understand the time-dependent spectral components of the developed fusion constructs. Further, the rate limiting step in the bioluminescent reaction involves the photoprotein reconstituting with CTZ; a factor that needs a grandiose consideration for long-term applications of bioluminescence in vivo[7,18].

4.6 Materials and Methods

4.6.1 Gene construction

All constructs were subcloned into pRSET-A (Invitrogen) and assembled by restriction enzyme cloning or using Gibson assembly methods (NEBuilder, HiFi DNA Assembly Kit, NEB) with custom designed oligos. Assemblies were roughly verified by restriction enzyme digest. Unless noted, blue (eCFP, mCerulean, mTFP1, mTagBFP2) and red-shifted (mScarlet, mRuby3, tdTomato) constructs were created by fusing the fluorescent protein's C-terminus and photoprotein's (OBE or AEQ) N-terminus by a standard linker provided by Bakayan et al (OSP #72764) (GTELYKSGGSGSGGQSG)[22]. Tandem monomer fluorescent proteins were creating by fusing the respective protein repeats by a 12-reside amino acid linker (GHGTGSTGSGSS) as described by Campbell et al[29]. Where described and labeled, fluorescent proteins from red-shifted fusion constructs were truncated; V1-V4 represents successive multiples of 2x amino acid truncations into the C-terminus of the labeled fluorescent protein (V1 is equivalent to the original protein, V4 is equivalent to a six amino acid C-terminal deletion). Truncated fluorescent proteins were fused to AEQ or OBE using amino acids (SG) or (GT) as a linker. All truncated red-shifted fusion proteins also utilized a shortened version of AEQ, truncated nine amino acids into the N-terminus (new amino acid sequence start FDNPRW)[21]. The gene encoding mScarlet was ordered as a G-block from IDT, the sequence was obtained from Bindels et al[27]. The gene encoding apo-obelin was ordered as a G-block from IDT, the sequence was obtain from Vysotski et al[8]. The gene encoding apo-aequorin-macrodactyla was ordered as a G-block from IDT, the sequence was obtain from Luo et al[30]. The gene coding for mTFP1[31] (# 55469), mRuby3[28] (#74252) were obtain from

addgene. The genes for eGFP-AEQ and tdTomato-AEQ were provided by Bakayan et al (OSP #72764)[22]. All other constructs were provided by the Cornell BRC imaging core. All transformed genetic constructs were stored as glycerol stocks in NEB 5-alpha or BL21 (DE3) competent E. Coli pending experimental needs.

4.6.2 Protein production, purification, and storage

Glycerol stocks coding for a series of FP-AEQ, FP-OBE, or FP-MAC fusions were streaked on ampicillin plates (100µg/mL) and allowed to grow for 14-18 hours at 37ºC. Log phase colonies of the desired indicator were selected as starter colonies and grown for 14-18 hours at 37ºC shaking at 225rpm in 5 mL Luria Bertani broth (100ug/mL ampicillin disodium salt (Sigma), 0.5% D-Glucose). After outgrowth, proteins were scaled 1:50 or 1:100 in 500mL a Erlenmeyer flask containing freshly autoclaved Terrific Broth and 100ug/mL Ampicillin. Shaking continued at 190rpm for 8 hours at 37ºC until OD of .6-.8 was reached. 1M IPTG was added to the solution (final concentration of 1mM), and induction was carry out for an additional 3 hours at 37ºC. After induction, the incubator was turned down to 12ºC to help induce protein expression and inhibit bacterial growth overnight. A small aliquot (500ul) was sampled from each flask and screened qualitatively under a transilluminator for the presence of the fluorescent portion of the photoprotein fusions. Once expression was induced, the bacteria were removed from shaker and stored at 4ºC for 24 hours. These 500mL outgrowths were then poured into di-electric 1L centrifuge bags and spun in a Beckman JLA9.1000 rotor at 6300xg, 4ºC, for 45minutes. Post centrifugation, remaining media was poured off and the pellet was resuspended in 25mL of 1xTris-EDTA in a 50mL

propylene column. The suspension was centrifuged 6000xg for 10min at 4C. The supernatant was decanted and the pellet was resuspended in 22mL of lysis buffer (50 mM NaH_2PO_4, 300mM NaCl, 3 mM imidazole 1ul/mL protease inhibitor cocktail solution (BioVision EZBlock Protease inhibitor cocktail, EDTA-free) was added fresh for each use). The resuspended solutions were stored at -80ºC to promote lysis at a future date.

Resuspended bacterial pellets were thawed, sonicated on ice (40% amplitude, 20 beats, 10 minute intervals, x2). The lysates were centrifuged at 13000xg for 10 minutes at 4ºC. Clarified supernatants were filter through a 0.8µm low protein binding filter using a 30mL syringe into a fresh 50mL propylene tube. 2mL of Ni-NTA resin slurry was centrifuged, and the clear ethanol supernatant was removed. Each protein was mixed with the remaining Ni-NTA beads, covered in aluminum foil, and allowed to rotate at 4C overnight. Ni-NTA bound proteins were loaded into reusable 5ml fritted propylene columns (Qiagen) until all flow through was collected. Protein bound resin was washed with wash buffer 3x10mL (50mM NaH2PO4H20, 300mM NaCl, 10mM imidazole, pH adjusted with NaOH to 7.97). Proteins were eluted with 2x2mL of 500mM imidazole buffer (50 mM NaH2PO4, 300mM NaCl, 500 mM imidazole, pH 8.0).

Millipore Ultracell (Ultra-4 Amicon 30kDA) buffer exchange columns were pre-equilibrated with 1xTE (Sigma 100x Tris-EDTA in ddH20) containing 1ul of 10% pluronic F-68 per 50mL of 1xTE (only for equilibration to prevent proteins from sticking to the filter membrane). Buffer exchange was performed using the 30kDA Ultracell membrane with 1xTE containing 5mM Beta-mercaptoethanol, this was

repeated x4 (spinning down to 100ul and resuspended in 4mL of the new buffer). 80% glycerol (in water) was added to approximately 750ul of resuspended protein to make a 5% storage additive, and brought up to 900 µl total, then sterile filtered through a .22 PVDF - sigma membrane to prevent bacterial growth. Samples were stored in 250µl aliquots by snap freezing in dry ice ethanol, then storing at -80C. For spectral measurements, the proteins were thawed, reconstituted with excess CTZ overnight (coelenterazine-H or coelenterazine-Native from Nanolight, cat#3011, cat#3031) , and buffer exchanged pending experimental needs.

4.6.3 Bioluminescent spectral measurements

For collection of bioluminescent spectra, proteins were concentrated between 1 and 10mM after coelenterazine reconstitution (absolute quantitative methods are difficult to obtain due to mixture of reconstitution components and changes in absorption pending state of calcium discharge). 50-100µL of concentrated protein was loaded into an optically clear cuvette and placed into a chamber containing a fiber optic coupler. A small fiber optic spectrometer (Ocean Optics QE65000, 1kHz sampling, 100ms integration, background subtracted) was used to measure the spectral output of each bioluminescent protein in response to saturating calcium (200mM CaCl2 in 100mM KCl, 10mM MOPS). Indicator peak wavelength are represented as a normalized plot from one integration window. For red-shifted indicators, average peak intensity values across all collected spectra were used to determine peak donor (OBE, AEQ, MAC) or acceptor FP wavelength, and calculate the ratio between the two peaks.

For determining relative luminescent output (RLO), reconstituted photoproteins

were resuspended in calcium free buffer (100mM KCl, 10mM MOPS, 0.5mM EGTA, pH 7.24) and brought into contact with calcium buffer via pipette (100mM KCl, 10mM MOPS, 10mM $CaCl_2$, pH 7.24). Light output is monitored constantly and calculated as mean gray value per frame collected with Basler AG CMOS camera (Basler acA2040-55uc); averaged over three trials. Light output is normalized to protein concentration as determined by absorbance at the FP-acceptor's peak wavelength, collected on a Cary UV-300 spectrophotometer. The total light integrated over the entire time course is stated as a (%) relative to eGFP-Aeq[22]. Experiments were run in parallel in 96-well plates to improve relative precision. Rapid mixing kinetic measurements were conducted on a stopped flow (SFM-400 Biologic) where protein bioluminescence in calcium-free buffer was measured in response to rapid mixing of saturating calcium solution (200mM $CaCl_2$, 10mM MOPS).

4.7 Supplemental Material

The spectra for certain fusion proteins created here or provided externally were characterized on our instrumentation as described in the main body of the text. Normalized absorbance spectra are plotted in **Figure 4.4**. Non-normalized absorbance spectra from the same dataset were utilized in RLO calculations.

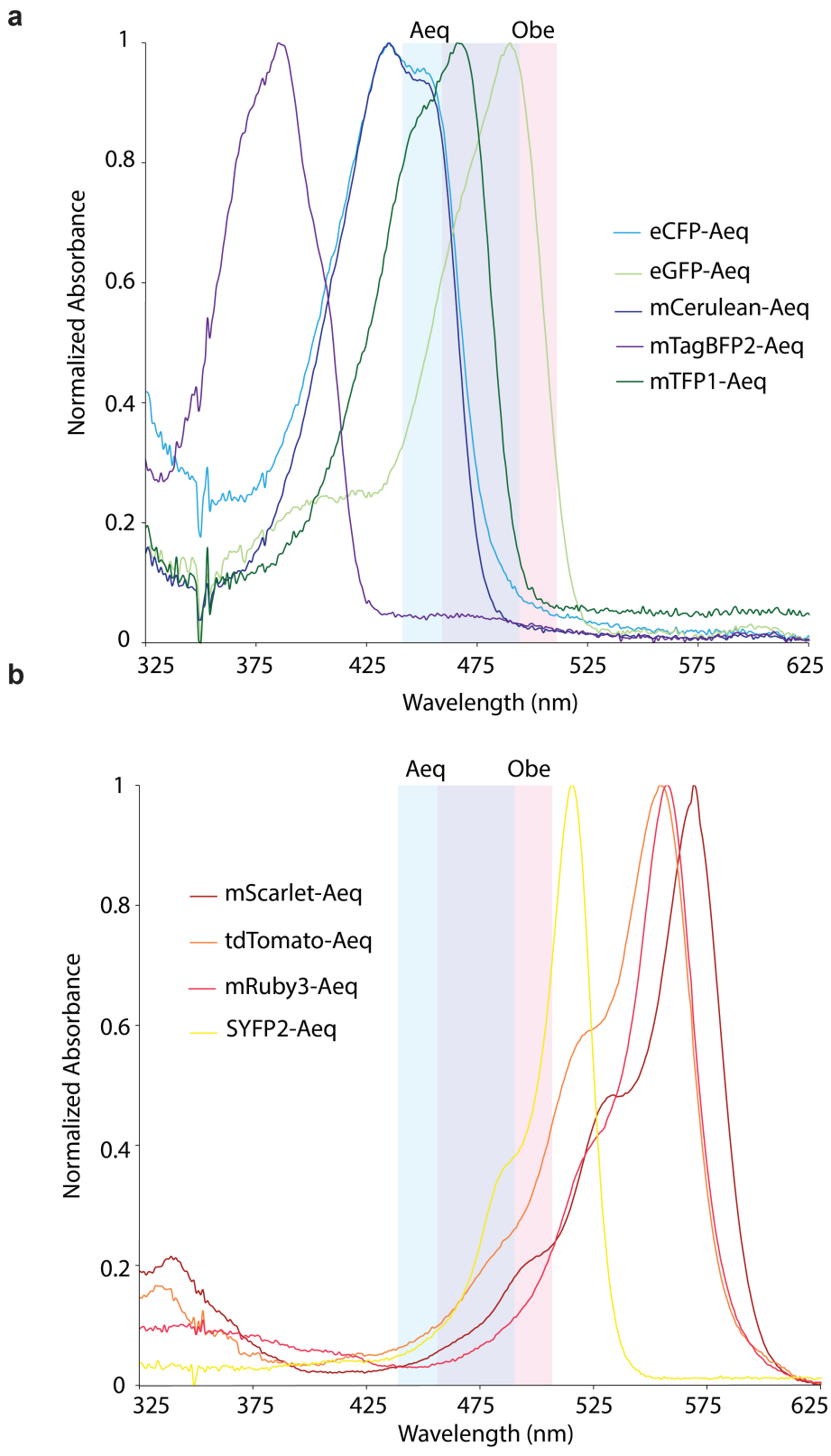


Figure 4.4 Absorbance profiles of blue and red fluorescent proteins
Absorbance of (**a**) blue and (**b**) red fluorescent proteins fused to AEQ. A FWHM profile of aequorin an obelin is plotted behind the absorbers to display partial donor emission overlap.

4.8 References

1. Frazier, H. N. *et al.* Calcium's role as nuanced modulator of cellular physiology in the brain. *Biochem. Biophys. Res. Commun.* **483,** 981–987 (2017).
2. Forostyak, O. *et al.* Physiology of Ca 2+ signalling in stem cells of different origins and differentiation stages. *Cell Calcium* **59,** 57–66 (2016).
3. Kotlikoff, M. I. Genetically encoded Ca2+indicators: Using genetics and molecular design to understand complex physiology. *J. Physiol.* **578,** 55–67 (2007).
4. Augustine, G. J., Santamaria, F. & Tanaka, K. Local calcium signaling in neurons. *Neuron* **40,** 331–346 (2003).
5. Burgoyne, R. D. & Haynes, L. P. Sense and specificity in neuronal calcium signalling. *Biochim. Biophys. Acta - Mol. Cell Res.* **1853,** 1921–1932 (2015).
6. Berridge, M. J., Bootman, M. D. & Roderick, H. L. Calcium signalling: Dynamics, homeostasis and remodelling. *Nat. Rev. Mol. Cell Biol.* **4,** 517–529 (2003).
7. Shimomura, O., Johnson, F. H. & Saiga, Y. Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusan,Aequorea. *J. Cell. Comp. Physiol.* **59,** 223–239 (1962).
8. Illarionov, B. A., Bondar, V. S., Illarionova, V. A. & Vysotski, E. S. Sequence of the cDNA encoding the Ca2+-activated photoprotein obelin from the hydroid polyp Obelia longissima. *Gene* **153,** 273–274 (1995).
9. Gealageas, R. *et al.* Bioluminescent properties of obelin and aequorin with novel coelenterazine analogues. *Anal. Bioanal. Chem.* **406,** 2695–2707 (2014).

10. Drobac, E., Tricoire, L., Chaffotte, A. F., Guiot, E. & Lambolez, B. Calcium imaging in single neurons from brain slices using bioluminescent reporters. *J. Neurosci. Res.* **88,** 695–711 (2010).

11. Stepanyuk, G. A. *et al.* Interchange of aequorin and obelin bioluminescence color is determined by substitution of one active site residue of each photoprotein. *FEBS Lett.* **579,** 1008–1014 (2005).

12. Malikova, N. P. *et al.* Spectral tuning of obelin bioluminescence by mutations of Trp92. *FEBS Lett.* **554,** 184–188 (2003).

13. Malikova, N. P., Borgdorff, A. J. & Vysotski, E. S. Semisynthetic photoprotein reporters for tracking fast Ca2+ transients. *Photochem. Photobiol. Sci.* **14,** 2213–2224 (2015).

14. Morse, D. & Tannous, B. A. A Water-Soluble Coelenterazine for Sensitive In Vivo Imaging of Coelenterate Luciferases. *Mol. Ther.* **20,** 692–693 (2012).

15. Bakayan, A., Domingo, B., Vaquero, C. F., Peyriéras, N. & Llopis, J. Fluorescent Protein–photoprotein Fusions and Their Applications in Calcium Imaging. *Photochem. Photobiol.* **93,** 448–465 (2017).

16. Belogurova, N. V. & Kudryasheva, N. S. Discharged photoprotein obelin: Fluorescence peculiarities. *J. Photochem. Photobiol. B Biol.* **101,** 103–108 (2010).

17. Lee, J. Perspectives on Bioluminescence Mechanisms. *Photochem. Photobiol.* **93,** 389–404 (2017).

18. SHIMOMURA, O. The discovery of aequorin and green fluorescent protein. *J. Microsc.* **217,** 3–15 (2005).

19. Gorokhovatsky, A. Y. *et al.* Fusion of Aequorea victoria GFP and aequorin provides their Ca 2+-induced interaction that results in red shift of GFP absorption and efficient bioluminescence energy transfer. *Biochem. Biophys. Res. Commun.* **320,** 703–711 (2004).

20. Titushin, M. S. *et al.* NMR-derived topology of a GFP-photoprotein energy transfer complex. *J. Biol. Chem.* **285,** 40891–40900 (2010).

21. Bakayan, A., Domingo, B., Miyawaki, A. & Llopis, J. Imaging Ca2+ activity in mammalian cells and zebrafish with a novel red-emitting aequorin variant. *Pflugers Arch. Eur. J. Physiol.* **467,** 2031–2042 (2015).

22. Bakayan, A., Vaquero, C. F., Picazo, F. & Llopis, J. Red fluorescent protein-aequorin fusions as improved bioluminescent Ca2+ reporters in single cells and mice. *PLoS One* **6,** (2011).

23. Baubet, V. *et al.* Chimeric green fluorescent protein-aequorin as bioluminescent Ca2+ reporters at the single-cell level. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 7260–5 (2000).

24. Curie, T., Rogers, K. L., Colasante, C. & Brûlet, P. Red-shifted aequorin-based bioluminescent reporters for in vivo imaging of Ca2+ signaling. *Mol. Imaging* **6,** 30–42 (2007).

25. Clegg, R. M. Chapter 1 Förster resonance energy transfer—FRET what is it, why do it, and how it's done. in 1–57 (2009). doi:10.1016/S0075-7535(08)00001-6

26. Arts, R., Aper, S. J. A. & Merkx, M. Engineering BRET-Sensor Proteins. *Methods Enzymol.* **589,** 87–114 (2017).

27. Bindels, D. S. *et al.* MScarlet: A bright monomeric red fluorescent protein for

cellular imaging. *Nat. Methods* **14,** 53–56 (2016).

28. Bajar, B. T. *et al.* Improving brightness and photostability of green and red fluorescent proteins for live cell imaging and FRET reporting. *Sci. Rep.* **6,** 20889 (2016).

29. Campbell, R. E. *et al.* A monomeric red fluorescent protein. *Proc. Natl. Acad. Sci.* **99,** 7877–7882 (2002).

30. Luo, W.X., Zhang, J., Yang, H.J., Li, S.W., Xie, X.Y., Qin, Y.X. Pang, S.Q., Li, S.J. and Xia, N. S. *Aequorea macrodactyla apoaequorin (Aequorin) gene, complete cds*. (2000).

31. Ai, H., Olenych, S. G., Wong, P., Davidson, M. W. & Campbell, R. E. Hue-shifted monomeric variants of Clavularia cyan fluorescent protein: identification of the molecular determinants of color and applications in fluorescence imaging. *BMC Biol.* **6,** 13 (2008).

CHAPTER 5

5: Fluorescent Protein mNeonGreen from *B. lanceolatum* Tolerates Circular Permutation and Integrates into Sensing Domains of Ca^{2+} or ATP/ADP Ratio

5.1 Abstract

Genetically encoded fluorescent protein indicators have evolved to meet many experimental imaging needs. One common strategy to engineer indicators has focused on circular permutation of genes primarily isolated from either *Aquorea Victoria* or *Anthozoa*, gene fusion to a desired sensing domain, and subsequent mutagenesis to screen for optimal sensor properties. Yet, few have explored the sensing utility of the brightest green fluorescent protein created to date, mNeonGreen (mNG) derived from *B. Lanceolatum*. Here, we explore various circular permutations (cp) of mNG which serve as dynamic fluorophore substitutes within known calcium or ATP/ADP sensing domains. When fused to calcium sensor complex calmodulin/M13, cp-variants of mNG showed notable pH-dependent responses to changes in calcium concentration. When inserted into the T-loop of Glnk1's A-protomer subunit from Perceval-HR, cp-mNG variants demonstrated a response to ATP/ADP ratio. The identified and characterized cp-mNG variants presented can serve as a protein-platform for high-throughput indicator engineering strategies. Further, dynamic characteristics amongst cp-variants demonstrate the importance of a single amino acid frameshift in circular permutation strategies of mNG.

5.2 Introduction

Development of genetic tools remains essential to drive advances in the process of scientific discovery. Fluorescent proteins, genetically encoded chromatic ensembles of highly conserved structure, have been engineered to report a dynamic range of photophysical properties when exposed to various biophysical conditions[1,2]. Initiated by Roger Tsien and colleagues over two decades ago, circular permutation (cp) and receptor insertion into green fluorescent protein (GFP) isolated from *Aequorea Victoria* sparked a revolution in the employment of fluorescent protein indicators for understanding biological systems with light[3,4]. To date, fluorescent proteins have evolved primarily from genes isolated from the jellyfish species *Aequorea Victoria* and the class of marine invertebrates *Anthozoa*[5]. Many scientists have exploited these proteins to reveal subcellular structures across the entire visible spectrum via tagging and high-resolution imaging. While others have fused them to various protein receptor counterparts to detect physiologically relevant cellular changes in voltage, calcium, ATP/ADP or NADH[6–8]. As experiments become increasingly dynamic, new indicators need to be continually developed and explored.

Genetically encoded fluorescent indicators can suffer from a variety of undesirable photophysical characteristics such as rapid photobleaching, chromophore photoconversion, pH sensitivity, or target ligand binding requirements that fall outside the ideal dynamic range. In order to explore solutions to some of these limitations, we developed a variety of cp-variants of the brightest and most photostable monomeric green fluorescent protein, mNeonGreen (mNG), engineered from an alternative species: *Branchiostoma lanceolatum*[9]. We chose the calcium sensitive receptor protein complex

Calmodulin (CaM)/M13 from the GCaMP series of indicators[10–12], and the ATP/ADP sensing trimer GlnKI from Perceval-HR to explore the tolerability of different cp-mNG variants to receptor insertion[6,13]. Here, we report the photophysical characteristics of cp-mNG variants fused to the described indicator domains and identify ideal candidates for further engineering as single molecule protein sensors. We describe in detail one variant, cp146-145, which displays acceptable calcium dependent photophysical properties in solution when cloned into CaM/M13. We also display characteristics of four cp-variants cloned in place of cp-mVenus from Perceval-HR-mCherry and suggest cp145-144 as a possible candidate for sensing ATP/ADP ratio in solution; with some characteristics still under investigation.

5.3 Results

5.3.1 Circular permutation of mNeonGreen

In the quest for a cp-variant of mNeonGreen to use as a building block for an improved calcium indicator, we followed the method of Kotlikoff et al. by creating multiple clones of cp-mNG using PCR amplification of a tandem-mNG template as described (methods)[14]. However, instead of measuring fluorescence of "naked" cp-mNG, we ligated all cp-variants directly into an RSET-CaM/M13 bacterial expression vector. Initial cp-mNeonGreen-Calmodulin/M13 (cp-mNGCaMP) constructs focused on narrow regions emulated from previously described circular permutations of GFP starting at amino acid (aa) 148 (**Supplementary Table 2**)[15]. 13 additional constructs were created, shifting 1-aa deeper into the protein complex, ending with residue 134 ("-n", truncations were not explored). Attempts were also made to screen at aa-160 and aa-

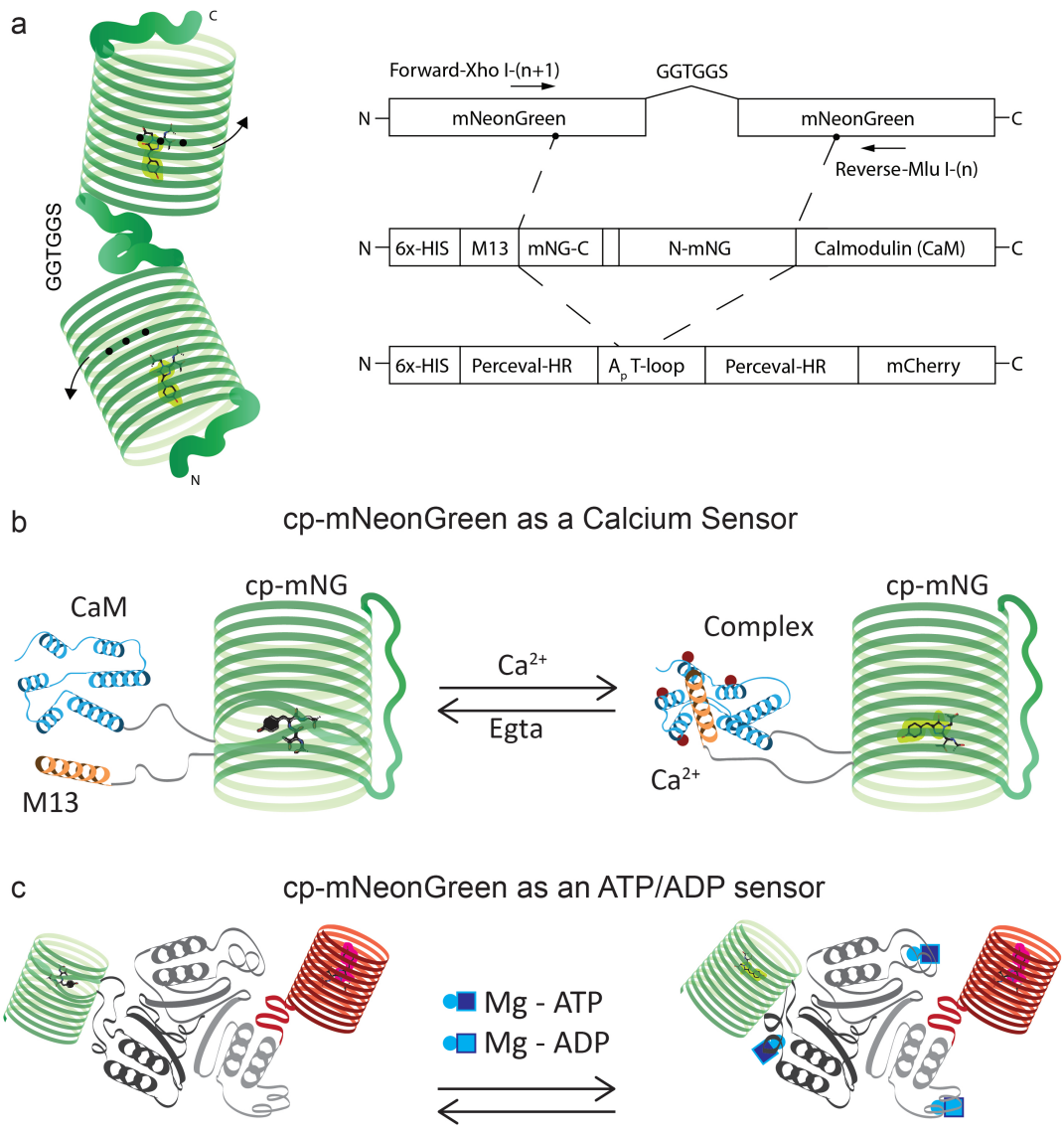


Figure 5.1 Cloning and mechanisms of cp-mNeonGreen constructs
a) Protein pictograph of a tandem mNG construct used as a template for cloning cp-variants. Arrows indicate priming direction and black dots indicate arbitrary circular permutation points. **b**) Schematic illustrating the anticipated mechanism of the CaM/M13 complex for calcium sensing with cp-mNG. In the absence of calcium cp-mNG may experience chromophore destabilization, leading to loss of fluorescence. In the presence of calcium elements of the fused complex may interact to stabilize the environment and return the fluorophore to a bright state. **c**) Schematic illustrating a possible mechanism of the GlnKI trimer for sensing Mg-ATP or Mg-ADP with cp-mNG. The geometry and/or conformational changes of the GlnKI trimer upon ligand binding may alter cp-mNG fluorophore function. Red-fluorescent protein mCherry is fused to the complex to aid in the normalization of any signal changes in cp-mNG variant testing.

170, creating 16 constructs in total. After plating and allowing for 18-hour incubation at 37°C, colonies were roughly screened for possible fluorescence using a transilluminator at 405nm (Invitrogen), and again following 24 hour incubation at 4°C. Only colonies displaying detectable fluorescence were chosen for large scale purification. **Figure 5.1a** demonstrates the overall cloning scheme for creating the cp-variants. The traditional soft (GGTGGS) linker was used to fuse two copies of mNG at their N and C termini as displayed[4]. Arrows indicate priming direction and black dots indicate circular permutation points which can be viewed in **Supplementary Table 2**. **Figure 5.1b** represents a schematic of the resultant clones of cp-mNG into CaM/M13 calcium sensitive binding domain, where the proposed mechanism of calcium sensing is illustrated. **Figure 5.1c** represents a schematic of the resultant clones of cp-mNG into GlnkI's T-loop from Perceval-HR[6], the proposed mechanism is illustrated.

5.3.2 Calcium-dependent spectroscopy

Although it is possible that all of the 16 constructs created may demonstrate calcium sensitivity, only a small portion were chosen for further testing. Figure **5.2a-d** show the results of characterization for (-c)-cp-mNGCaMP$^{146\text{-}145}$; which is referenced as "(-c)" to represent the variant label (**Supplementary Table 2**) and "NGCaMP" to suggest that it is designed to display calcium activity from interactions with the CaM/M13 complex. The change in fluorescence intensity upon binding calcium is calculated as in Eq. (1),

$$F_{int} = \frac{F - F_{min}}{F_{max} - F_{min}} \qquad \text{(Eq.1)}$$

where F_{min} represents the sum of emission signal intensity at 0.00μM [Ca^{2+}], F_{max} represents the sum of emission signal intensity at 39.8μM [Ca^{2+}], and F represents the

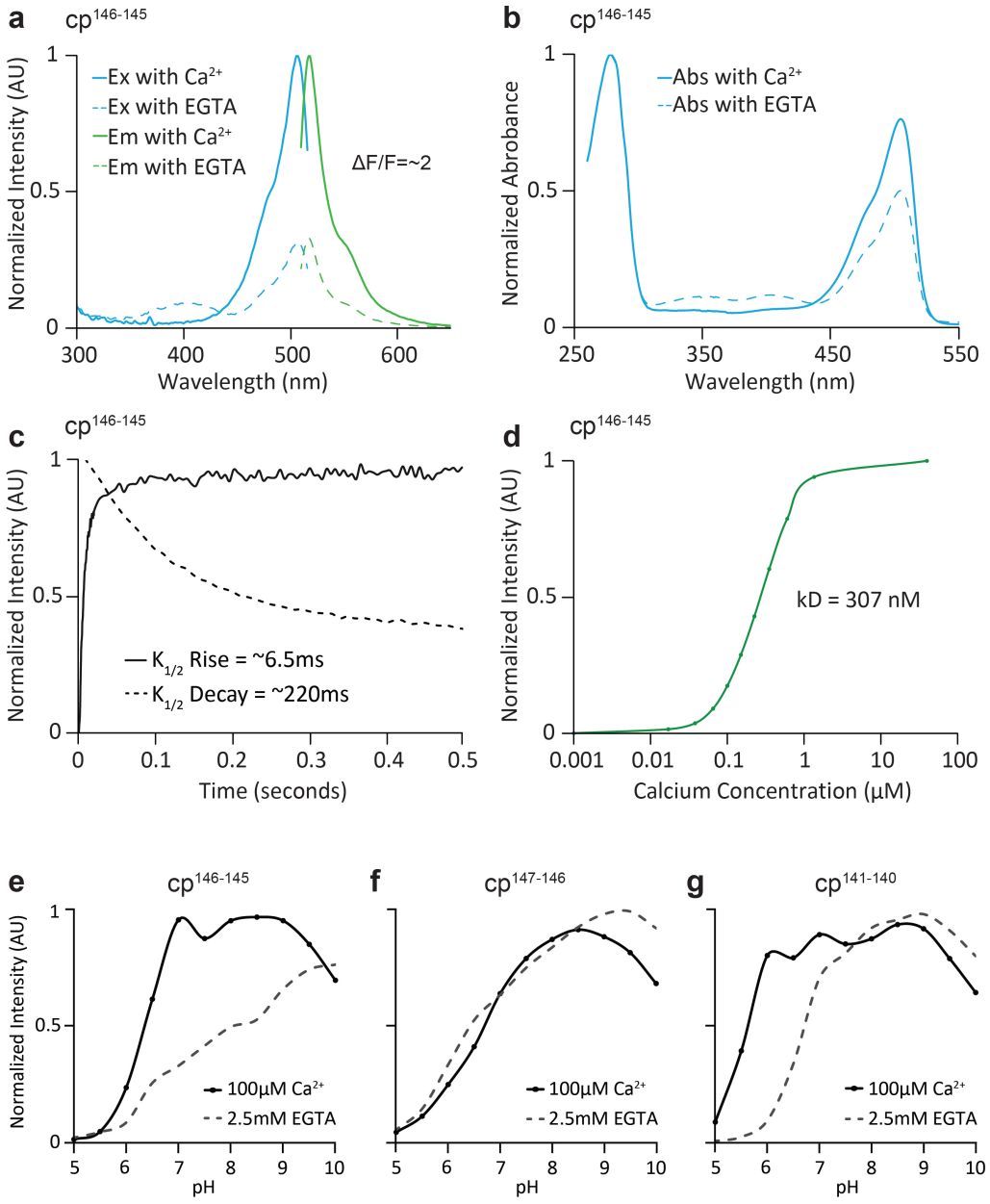


Figure 5.2 Spectroscopy of cp-mNG variants in CaM/M13 sensing domain
(**a**) Normalized excitation and emission spectra of cp146-145 with either 100μM $CaCl_2$ or 2.5mM EGTA (10mM HEPES pH 7.2). (**b**) Absorbance spectra with either 200μM $CaCl_2$ or 1.0mM EGTA (10mM MOPS buffer pH 7.8). (**c**) Normalized Kintek stopped-flow association/dissociation kinetics of indicator in response to (decay) 5mM EGTA or (rise) 5mM $CaCl_2$ (10mM MOPS buffer pH 7.8). (**d**) Logarithmic plot of normalized fluorescence as a function of free calcium concentration [μM] (Promokine K_2EGTA/CaEGTA). (**e**, **f**, **g**) pH-dependent fluorescence emission of cp-mNG variants cloned into CaM/M13 with either 100μM $CaCl_2$ or 2.5mM EGTA in respective pH 5-10 Carmody buffers. (**e**) variant cp146-145 as described in **a**-**d**, (**f**) variant cp147-146, and (**g**) variant cp141-140.

sum of emission intensities at a specified concentration along the curve shown in **Figure 5.2d**; emission is collected at excitation wavelength of 485nm. Utilizing the linear region of the curve, the K_D is calculated to be approximately 307nM (**Supplementary Figure 5.7**). Calcium dissociation kinetics were determined on a Kintek stopped-flow mixer in response to saturating $CaCl_2$ (K_{rise}) or EGTA (K_{decay}). Preliminary results suggest $K_{1/2\ rise}$ time of approximately 6.5 milliseconds and $K_{1/2\ decay}$ of approximately 220 milliseconds (**Fig 5.2c**).

5.3.3 pH-dependent spectroscopy of calcium indicators

As with most physiological indicators based on circularly permutated proteins, solvent exposure to the chromophore can alter its biophysical characteristics as fluorescent intensity values tend to rely on solvent pH[2]. To analyze this phenomenon in further detail for cp-mNGCaMP variants, we tested fluorescent intensity profiles in response to a range of pH 5-10 Carmody buffers containing 100μM $CaCl_2$ or 2.5mM EGTA[16]. The resulting fluorescent profiles for (-c)-cp-mNGCaMP$^{146-145}$, (-b)-cp-mNGCaMP$^{147-146}$, and (-h)-cp-mNGCaMP$^{141-140}$ are shown in **Figure 5.2 e-g** respectively. While all indicators shown are pH sensitive under calcium and calcium-free conditions (EGTA), they display unique characteristics. (-c)-cp-mNGCaMP$^{146-145}$ shows notable intensity changes at physiological pH (~7), which correspond with our calculations in **Figure 5.2a** and represents the best reporter of calcium that we have tested in this range. (-b)-cp-mNGCaMP$^{147-146}$ shows parallel dependence on calcium as a function of pH, deviating above pH 8, which could suggest its utility as a fusion partner around physiological pH. (-h)-cp-mNGCaMP$^{141-140}$ shows larger calcium dependent differences

at more acidic pH (~pH 6) and may be suitable as a calcium indicator where cells operate under more acidic extremes such certain types of cancer. Further, these pH dependent properties merit exploration of structural changes that occur at these different permutation sites. **Supplementary Figure 5.5** presents more detail concerning the excitation spectra of (-c)-cp-mNGCaMP$^{146\text{-}145}$ and (-b)-cpmNGCaMP$^{147\text{-}146}$, suggesting interesting properties between protonation states of the different indicators. Further, to derive a more precise comparison, the original plots from **Figure 5.2e**, **5.2f** and **5.2g** are displayed non-normalized in **Supplementary Figure 5.6**.

5.3.4 ATP/ADP Mg^{2+} dependent absorbance

Since different versions of cp-mNG report calcium sensitivity when inserted into sensing domain CaM/M13, we hypothesized that it may also be useful as a substitute for other indicators based on circularly permutated proteins. ATP/ADP ratiometric indicator Perceval-HR utilizes circularly permutated mVenus as its fluorescent reporter[6,13]. In recent studies, mVenus has been shown to have undesirable photobleaching properties, and its suitability as a circularly permutated indicator merits further review[17]. For these reasons, we subcloned four cp-mNG variants a-d (**Supplementary Table 2**) in place of cp-mVenus in Perceval-HR's GlnKI T-loop. In this application, the entire construct is also fused to mCherry to assist in normalization of ATP/ADP dependent responses. **Figure 5.3a-d** shows the respective absorbance profiles of the four different cp-mNG variants subcloned in place of mVenus normalized to the absorbance of mCherry at 586nm. While changes are notably small across all variants given ligand concentrations, careful pH control suggests cp-mNG

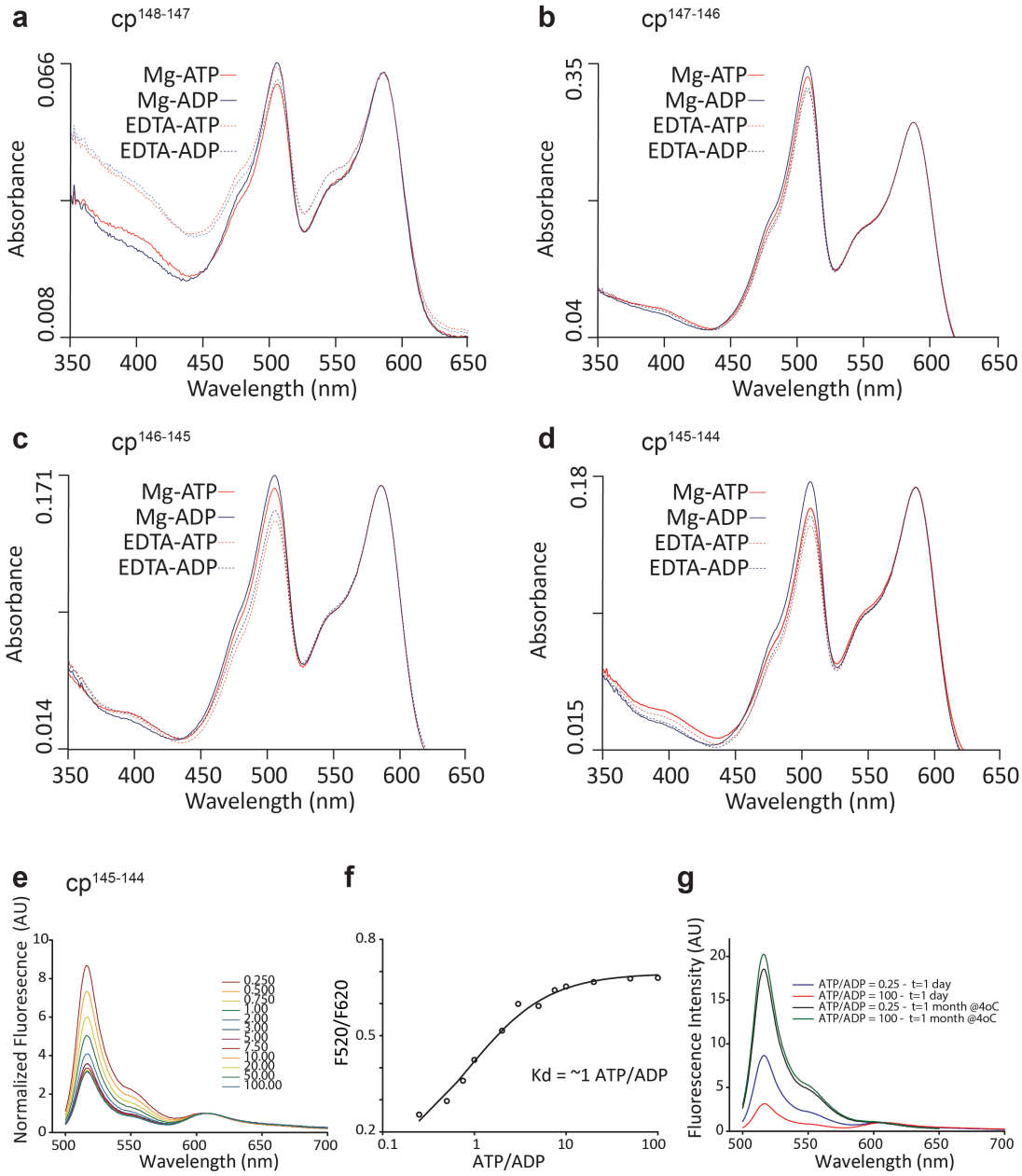


Figure 5.3 Spectroscopy of cp-mNG variants in Perceval-HR-mCherry sensing domain
Absorbance spectra of four different variants of cp-mNG ligated in place of cp-mVenus from Perceval-HR-mCherry. **a**) $cp^{148-147}$, **b**) $cp^{147-146}$, **c**) $cp^{146-145}$, **d**) $cp^{145-144}$. All spectra are background subtracted and normalized to mCherry's 586nm absorption peak. Legend (pH of all solutions adjusted to 7.29 +/- .02 in 10mM HEPES buffer): Mg-ATP (5mM $MgCl_2$, 1mM ATP); Mg-ADP (5mM $MgCl_2$, 1mM ADP); EDTA-ATP (1mM EDTA, 1mM ATP); EDTA-ADP (1mM EDTA, 1mM ADP). (**e**)Variant $cp^{145-144}$ shown in **d** above. Fluorescence emission normalized to mCherry (586nm) in response to ATP/ADP ratio. (**f**) Log plot of peak intensity ratio as collected in (**e**). (**g**) Comparative fluorescence emission after a period of storage at 4°C; notable intensity changes merits further investigation of this variant as a stable reporter of ATP/ADP ratio.

variants could be useful as a reporter of ATP/ADP ratio in this setting.

5.3.5 ATP/ADP Mg^{2+} dependent fluorescence

As mentioned earlier, physiological indicators based on circularly permutated proteins can display variable characteristics in response to chromophore environmental changes. To further examine the response profile of variant (-d)-cpMNG-Perceval-HR-mCherry$^{145\text{-}144}$, we measured its fluorescent emission across a range of ATP/ADP ratio concentrations. Upon initial testing, this variant showed impressive response profiles as a ratiometric indicator as seen in **Figure 5.3e** when normalized to the mCherry emission peak (586nm). However, upon long term storage at 4ºC, results were not reproducible (**Fig 5.3g**). Experimentation is ongoing to determine the cause of these results and confirm the validity of this variant.

5.3.6 ATP/ADP ratio response measurements in cultured MDA cells

As discussed earlier, physiological indicators are often influenced by their surrounding environment both in vitro and in vivo. To further investigate the properties of (-d)-cp-mNG-Perceval-HR-mCherry$^{145\text{-}144}$, we utilized lentiviral transduction techniques to generate and sort an MDA cancer cell line to understand how the indicator may report physiological changes in vivo. **Figure 5.4a** shows fluorescent images of MDA cells expressing this variant after lentiviral transduction and fluorescent sorting of the cell line. Initially, it is noted that the acidic lysosomal environment shows small bright red image pixels which could indicate protein degradation and pH sensitivity of the fusion pair. Upon treatment with 2-deoxyglucose (DG) and 10uM Oligo for 1 hour,

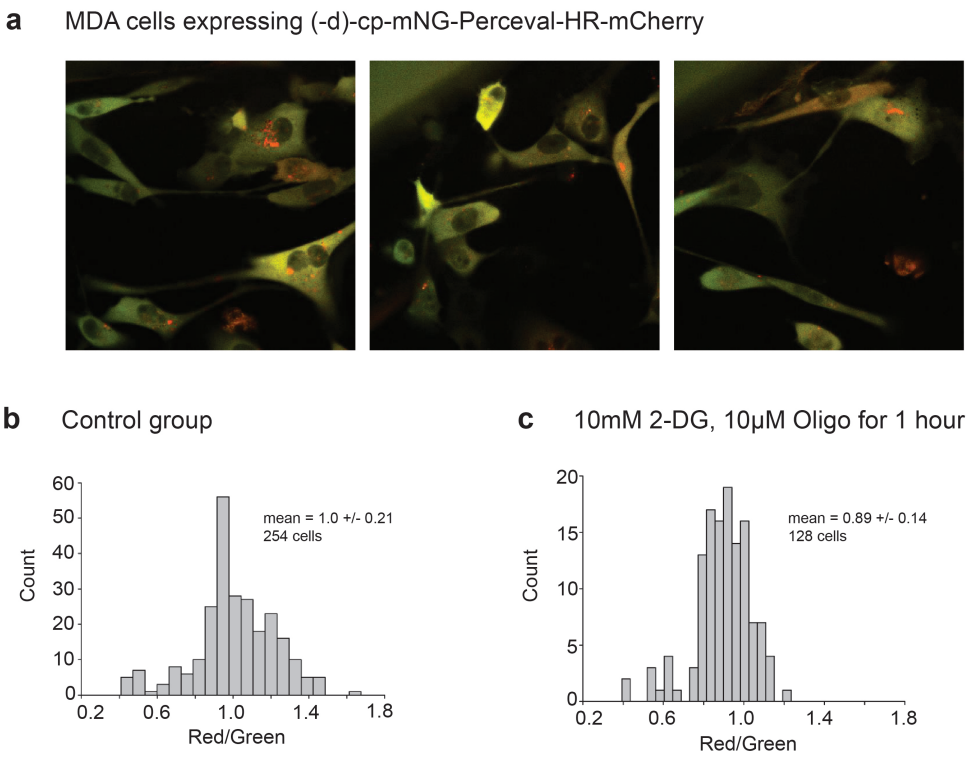


Figure 5.4 (-d)-cp-mNG-Perceval-HR-mCherry in MDA cells
a) Fluorescent images of the indicator in MDA cells. **b**) Ratio distribution of green and red imaging channels from defined cells under control conditions. **c**) Ratio distribution of green and red imaging channels from defined cells after treatment with 10mM 2-DG and 10µM Oligo for 1 hour. Mean and standard deviation included on each plot.

cells show an approximated 10% change in ratio reported between green and red imaging channels over many defined cells (**Fig. 4c**) vs control conditions (**Fig. 4b**). If metabolic inhibition is occurring from treatment, this change could help explain the in vitro results discussed earlier. Experiments are ongoing to verify fluorescent changes and directionality of the new indicator in response to its environment.

5.4 Discussion

5.4.1 mNeonGreen tolerates circular permutation

mNeonGreen tolerates circular permutation across a range of different locations and performed well as an indicator considering the only modification was the choice of permutation site. While indicator characterization was not exhaustively explored; most cp-variants displayed fluorescence when cloned directly into the CaM/M13 indicator domain (**Supplementary Table 2**). Of the indicators that were characterized, a single amino acid frameshift resulted in substantially different properties. One notable feature is the relative difference in dual band excitation profiles at 400nm and 506nm in **Supplementary Figure 5.5**. Recent studies suggest a complex pH dependence of mNG's chromophore environment when compared to *Aequorea* derived fluorescent proteins[18]. Although beyond the scope of this work, future exploration of the structural changes resulting from mNG circular permutation may reveal interesting properties in its role as an indicator. The structural components of mNG that give rise to its robust and bright fluorescence at physiological pH may be a considerable limitation and/or opportune target in designing future indicators based on cp-mNG[19].

5.4.2 (-c)-cp-mNGCaMP$^{146-145}$ compares with existing calcium indicators

To date, few have explored the utility of mNG as an indicator of calcium activity. In 2016, Barykina et al. explored mNG as a single molecule calcium indicator utilizing the troponin C binding domain, creating a split mNG at the same site as our optimal calcium sensor, residues 146-145. After extensive linker optimization and mutational screening, their efforts resulted in an indicator with a calcium sensitivity, K_D in the ~100nM range and a fluorescence decrease of approximately 2-fold upon calcium binding. The motivation for the indicator is presented as a having a smaller genetic payload and one less calcium binding site than Cam/M13 based indicators[20]. This is an interesting finding, as the indicator properties are similar to the results reported here, suggesting that this site of perturbation may be worth further exploration.

5.4.3 (-d)-cp-mNG-Perceval-HR-mCherry is a candidate for cell testing

The spectroscopic results presented earlier suggest issues of repeatability for cp-mNG as an indicator of ATP/ADP ratio. These results could be explained by protein degradation over long term storage, or by pH related issues in the respective testing buffer. While experiments are ongoing to determine the cause of these changes, the indicator still reports a change to some degree in non-reproducible fluorescent profiles, and in pH-controlled absorption spectra. These factors merit continued testing in cell culture as protein turnover in cells may influence issues related to protein degradation in solution. Results show promise for these constructs as a reporter of ATP/ADP ratio in our initial explorations of a lentiviral transduced MDA cancer cell lines.

5.5 Conclusions and Future Directions

Fluorescent protein mNeonGreen from *B. Lanceolatum* tolerates circular permutation and receptor insertion into sensing domains of ATP/ADP and Ca2+. The cp-variants described here suggest that a single amino acid frameshift can alter indicator response profile. Indicators (-c)-cp-mNGCaMP$^{146\text{-}145}$ and (-d)-cp-mNG-Perceval-mCherry$^{145\text{-}144}$ have biophysical characteristics that merit further testing in cell culture for calcium and ATP/ADP response, respectively. Since indicators are derived solely from circular permutation, high-throughput mutagenic screening would be an ideal future step for producing optimal indicators.

5.6 Materials and Methods

5.6.1 Generation of tandem mNeonGreen template

Primers were designed to create a tandem fusion template of mNG with a soft GGTGGS linker in between two mNG repeats via *in situ* digestion with Kpn-I and ligation with T4 DNA ligase. To create circularly permutated inserts, primers to amplify 5'-Xho-I forward and 3'-Mlu-I reverse (**Supplement section 5.7**) at desired locations along the tandem dimer were designed for PCR amplification (**Fig. 5.1a**)[14].

5.6.2 Creation of calcium sensitive mNeonGreen constructs

A range of cp-mNG variants were cloned by restriction enzyme digest with Xho-I and Mlu-I followed by T4 ligation into a bacterial expression vector containing CaM/M13 DNA sequences chosen from calcium indicator GCaMP8 due to availability (provided by Kotlikoff lab). Ligations were transformed into NEB BL21 (DE3) competent cells

per manufacturers protocol and cultured on LB/Agar plates containing carbenicillin (100µg/mL). Bacterial colonies of variants that produced notable fluorescent colonies under transillumination at 405nm were selected and grown in LB/Carbenicillin media and stored as glycerol stocks for downstream analysis. Primer sequences are available in the supplemental section.

5.6.3 Creation of ratio-metric ATP/ADP mNeonGreen constructs

NEB Gibson assembly was used to facilitate the cloning of the desired cp-mNG variant into the T-loop of Perceval-HR-mCherry's A-protomer subunit in place of mVenus (**Fig. 1c**). Perceval-HR-mCherry was provided in the bacterial expression vector pHisParallel from the Lammerding lab. DNA amplicons were assembled following manufacturer's protocol, transformed into NEB5-α competent cells, and cultured on LB/Agar plates containing carbenicillin (100µg/mL). Successful clones were verified by restriction digest and transformed into Bl21 (DE3) for downstream protein expression. Primer sequences are available in the supplement section.

5.6.4 Protein expression and purification

Glycerol stocks (25%glycerol, 75%LB/bacteria) of BL21(DE3) cells containing the gene of interest were streaked on fresh LB/carbenicillin (100µg/mL) plates and incubated at 37ºC for 16-24 hours. Isolated colonies were grown in overnight seed cultures and diluted 1:60 in .5-1L of LB/Carbenicillin (100µg/mL) media and grown at 37ºC until opaque (16-20hrs; OD ≈ 0.60-0.80); 1mM IPTG was added and incubated for 3 hours at 37ºC followed by 16-24 hours at 16ºC. Bacteria were stored at 4ºC

overnight or harvested immediately at 6300rpm, for 45 minutes at 4ºC (Beckman JLA 9.1000 rotor). Cultures were resuspended in lysis buffer (50mM NaH_2PO_4, 300mM NaCl, 3mM imidazole, pH 8.0) with fresh protease inhibitor cocktail tablet (1 tablet/200mL) and stored at -80C for an arbitrary time between 1-7 days to initiate lysis. Frozen cultures were thawed and sonicated on ice (40% amplitude, 40 beats, 0.3sec, repeated x2, Branson). The lysates were clarified at 15,000xg for 30min at 4ºC and filtered through a 1.2µm filter (GE Whatman). Clarified lysates were mixed with 2-3mL of Ni-NTA affinity resin (Qiagen) for 3-16 hours on a rotating platform at 4ºC. The resin was poured into a 15mL fritted propylene column (Qiagen), washed three times with 12mL of wash buffer (50mM NaH_2PO_4, 300mM NaCl, 20mM Imidazole, pH 8.0), and eluted with 2x2mL of elution buffer (50mM NaH2PO4, 300mM NaCl, 500mM Imidazole, pH 8.2). Combined eluates were desalted and concentrated using Amicon Ultra-15 (30k or 50k MWCO) with 10mM MOPS/100mM KCl (pH 7.8) 3-5 times. Proteins concentrated at 5-30mg/mL were stored at 4ºC for immediate testing or flash frozen in liquid nitrogen and stored long-term at -80ºC. Proteins were roughly quantified by 280nm absorbance using a NanoDrop1000.

5.6.5 Spectroscopy of calcium indicators

The excitation and emission profiles of purified 6xHis-tagged proteins were diluted 50 fold in buffer (20mM HEPES, pH7.2) containing either 100µM CaCl2 or 2.5mM EGTA and analyzed using a Photon Technologies International (PTI) fluorometer outfitted with Windows 98 and internal software for spectral correction. The excitation profiles were scanned from 276nm-516nm (2nm step size, 0.4 second integration) against an

emission set at 530nm. The emission profiles were scanned from 500nm-650nm (2nm step size, 0.4 second integration) against an excitation set at 485nm. System gain and manual slit width were kept the same when possible. To collect absorbance profiles, proteins were diluted between 0.2-2mg/mL in storage buffer (10mM MOPS, 100mM KCl, ph7.8) and scanned across the visible spectrum with a Cary 300 UV-Vis Bio spectrophotometer. Calcium affinity curves were collected by serial exchange of dilute protein in K_2EGTA/CaEGTA buffers (PromoKine) following manufacturers protocol; spectra were collected using the PTI fluorimeter with identical emission scan settings mentioned above. To probe calcium dependent pKa, a series of Carmody buffers from pH 5-10 were prepared with either 100µM $CaCl_2$ or 2.5mM EGTA and scanned as describe above, or in replicate 96-well plate format using Zeiss 710 confocal (5x air objective, 4 frame average, 488nm excitation, 500-600nm emission).

5.6.6 Spectroscopy of ATP/ADP indicators

All ATP/ADP ratiometric indicators absorbance spectra were collected on a Cary 300 UV-Vis Bio spectrophotometer, background subtracted, and normalized to mCherry's 586nm absorption peak. Solutions for absorption measurements are reported as free concentrations of respective chelator, pH of all solutions adjusted to 7.29 +/- .02 in 10mM HEPES buffer: Mg-ATP (5mM $MgCl_2$, 1mM ATP); Mg-ADP (5mM $MgCl_2$, 1mM ADP); EDTA-ATP (1mM EDTA, 1mM ATP); EDTA-ADP (1mM EDTA, 1mM ADP). The spectral properties of proteins were analyzed using a PTI fluorometer. Solutions for ratiometric analysis were created from mixing the following stocks at respective volumes: Mg-ATP (30mM ATP, 30mM $MgCl_2$ in 30 mM HEPES pH 7.4)

and ADP (30 mM ADP in 30 mM HEPES pH 7.4). The emission profiles were scanned from 500nm-650nm (2nm step size, 0.4 second integration) against an excitation set at 485nm.

5.6.7 Lentiviral generation and cell transduction

All plasmids for lentiviral transfection were provided by the Lammerding lab. The lentiviral vector plasmid pCDH harboring a blasticidin resistance marker and was used for cloning constructs for mammalian transfection. For (-d)-cp-mNG-Perceval-HR-mCherry, restriction enzyme cloning was used to transfer the bacterial expression gene into the pCDH mammalian lentiviral expression cassette using the following primers:

F-NheI-Mamm-Perc atctaagctagcgccaccatggatccgaaa

R-FseI-PercmChr TAGTATGGCCGGCCttatctagatccggtg

Successful clones were screened using restriction enzyme digest and gel electrophoresis. Lentiviral constructs were produced in a HEK293T producer cell line in DMEM (11995-065 Invitrogen), 10% FBS, 1%P/S, 1% Glutamax at 5% CO_2 and 37.0^o C. Cells were passaged into a T-75 flask (Corning) to be ~70% confluent at the time of transfection. Cells were triple transfected using a PureFection transfection reagent (SBI) following manufacturer's instructions with the following plasmids in a 1:1:1 molar ratio: pSPAX2.GAG (1.55µg/mL – **4.5µL**); pMDG2.ENV (0.753µg/mL – **5.1µL**); pCDH-(-d)-Perceval-mCherry (0.65µg/mL – **10uL**). Media was freshly changed to include the transfection mix, and cells were allowed to incubate for 60 hours at 5% CO_2 and 37.0^o C. Media was carefully collected and concentrated utilizing Lenti-X concentrator according to manufacturer's protocol. The resulting pellet was resuspended

in sterile PBS (800µL), aliquoted, and immediately stored at -80ºC (no filtration). Freshly thawed MDA cells at 40% confluence in a T-25 flask were transduced by adding a 200µL aliquot (25% of total virus collected) of the concentrated lentivirus in 4mL of fresh media (no polybrene used). Cells were passaged twice, and fluorescently sorted for moderate brightness. They were plated on MatTek glass bottom dishes for imaging experiments. Experiments are ongoing at the time of writing this dissertation.

5.7 Supplemental Material

Table 2: cp-mNG construct labels and circular permutation locations

Sample ID	CP site	AA Resiudes	%Intensity relative to (-c)
(a)	148-147	W-D	70%
(-b)	147-146	D-A	95%
(-c)	146-145	A-A	100%
(-d)	145-144	A-T	45%
(-e)	144-143	T-L	18%
(-f)	143-142	L-S	20%
(-g)	142-141	S-N	49%
(-h)	141-140	N-T	20%
(-i)	140-139	T-M	77%
(-j)	139-138	M-V	28%
(-k)	138-137	V-P	3%
(-l)	137-136	P-G	0%
(-m)	136-135	G-D	14%
(-n)	135-134	D-A	9%
(10+fb)	160-159	T-K	N/A
(20+fb)	170-169	Y-T	N/A

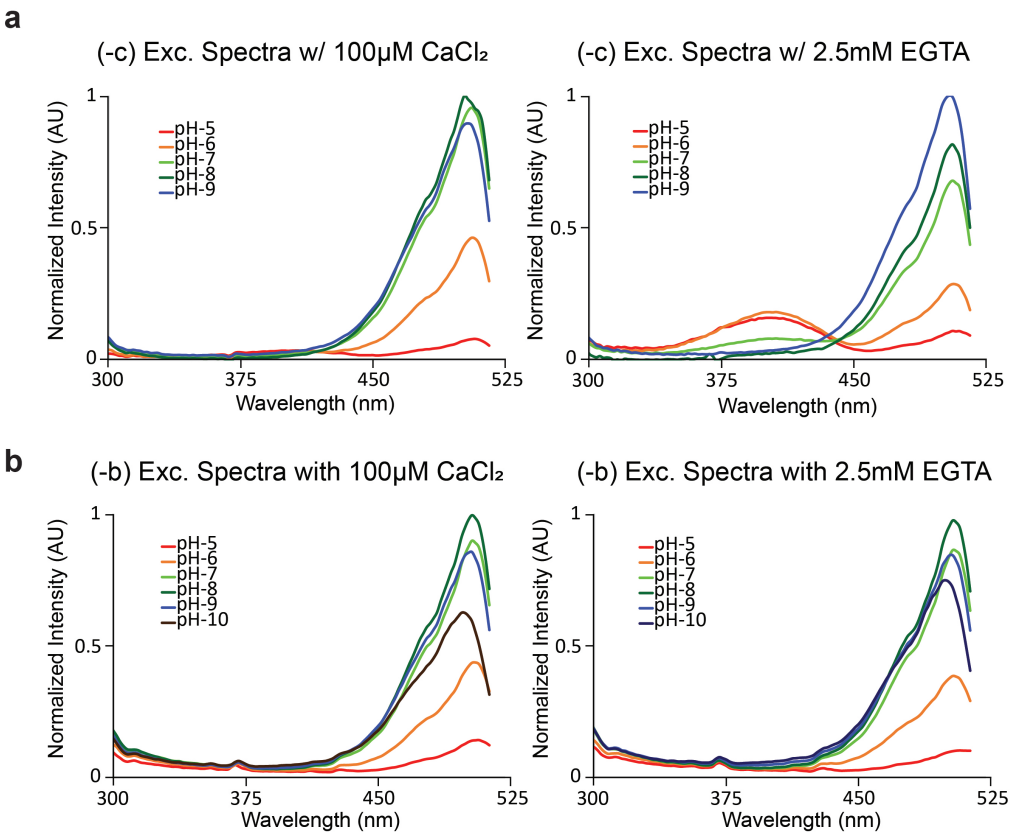


Figure 5.5 pH dependent fluorescence comparison of (-c) and (-b) cp-mNGCaMP
a) Normalized fluorescence excitation spectrum of (-c)-cp-mNGCaMP (emission 530nm) in ph5-9 Carmody buffer with 100 µM $CaCl_2$ or 2.5mM EGTA. **b**) Normalized fluorescence excitation spectrum of (-b)-cp-mNGCaMP (emission 530nm) in ph5-10 Carmody buffer with 100 µM $CaCl_2$ or 2.5mM EGTA.

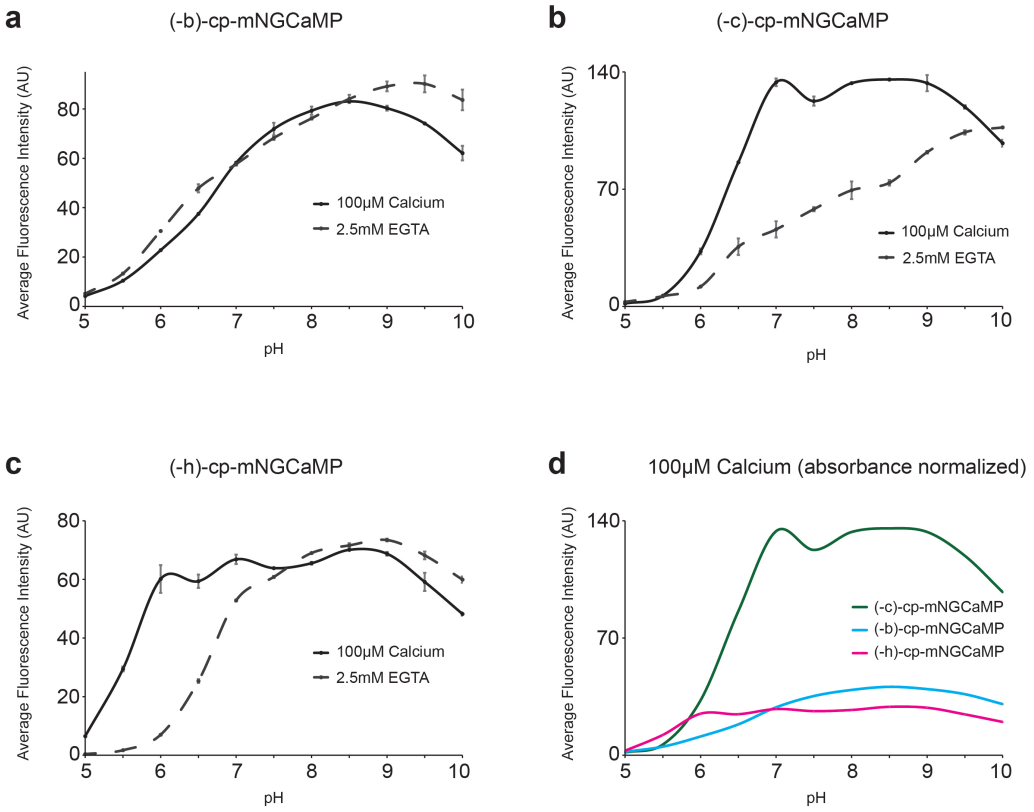
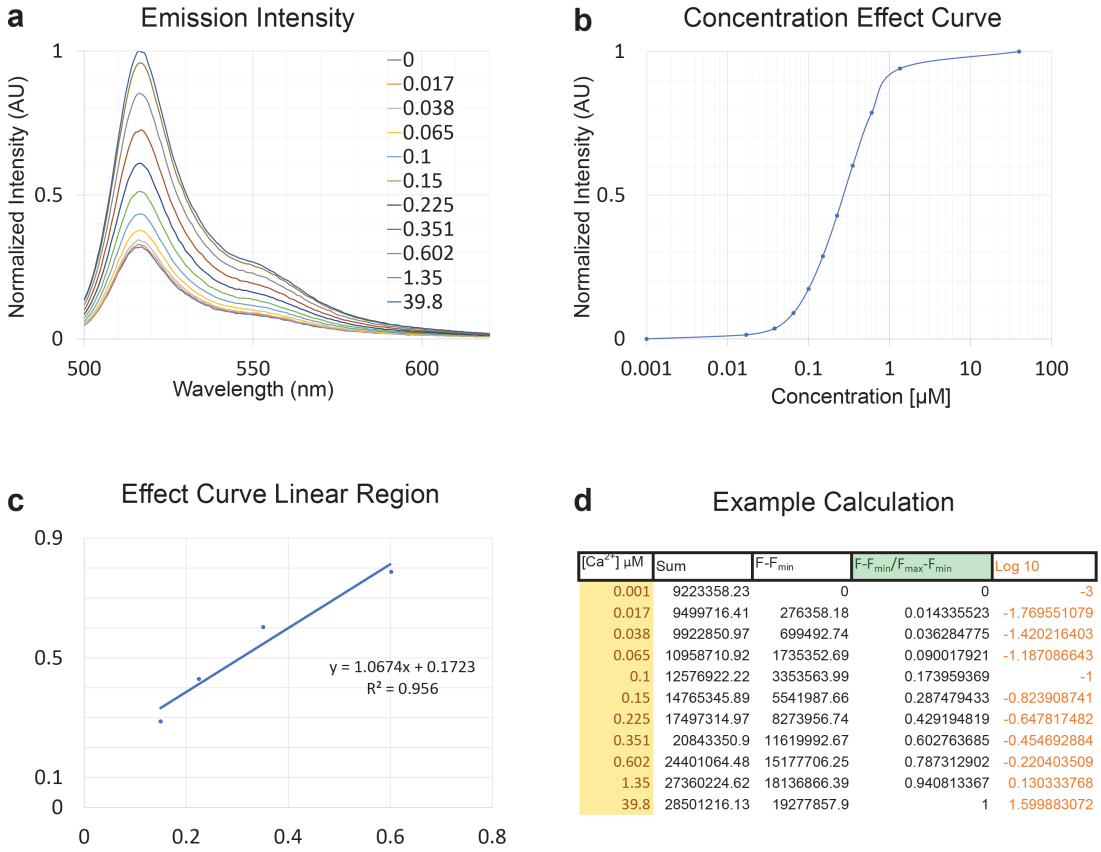


Figure 5.6 pH dependent fluorescence comparison of (-b), (-c), and (-h) cp-mNGCaMP
a) Averaged fluorescence emission spectrum of (-b)-cp-mNGCaMP in ph5-10 Carmody buffer with 100 µM $CaCl_2$ or 2.5mM EGTA. **b**) Averaged fluorescence emission spectrum of (-c)-cp-mNGCaMP in ph5-10 Carmody buffer with 100 µM $CaCl_2$ or 2.5mM EGTA. **c**) Averaged fluorescence emission spectrum of (-h)-cp-mNGCaMP in ph5-10 Carmody buffer with 100 µM $CaCl_2$ or 2.5mM EGTA. **d**) Overlay of all three indicators normalized to absorbance at 490nm for comparison of brightness. All results are averaged from n=3 trials, SD error bars shown in **a-c**.



[Ca^{2+}] µM	Sum	F-F$_{min}$	F-F$_{min}$/F$_{max}$-F$_{min}$	Log 10
0.001	9223358.23	0	0	-3
0.017	9499716.41	276358.18	0.014335523	-1.769551079
0.038	9922850.97	699492.74	0.036284775	-1.420216403
0.065	10958710.92	1735352.69	0.090017921	-1.187086643
0.1	12576922.22	3353563.99	0.173959369	-1
0.15	14765345.89	5541987.66	0.287479433	-0.823908741
0.225	17497314.97	8273956.74	0.429194819	-0.647817482
0.351	20843350.9	11619992.67	0.602763685	-0.454692884
0.602	24401064.48	15177706.25	0.787312902	-0.220403509
1.35	27360224.62	18136866.39	0.940813367	0.130333768
39.8	28501216.13	19277857.9	1	1.599883072

Figure 5.7 K_D calculation for (-c)-cp-mNGCaMP from calcium calibration kit
a) Fluorescence intensity plotted as a function of free calcium concentration in [µM] (PromoKine K_2EGTA/CaEGTA). **b**) Normalized log plot of integrated intensity as represented in **a**; F-F_{min}/F_{max}-F_{min} vs. calcium concentration. **c**) Plot of the linear region in **b** used to calculate K_D at ½ F_{max}. **d**) Example calculations for generating plot in **b** from data in **a**.

5.7.1 Primers used for cp-mNG construct generation

(a)CP:
XhoI_aCP1_F ccagCTCGAGtggtgcaggtcgaaga
MluI_aCP1_R gatACGCGTgtccgcagcggtca

(-b)CP:
(-)bCP-F ccagCTCGAGgactggtgcaggtc
(-)bCP-R agatACGCGTcgcagcggtcagc

(-c)CP:
(-)cCP-F ccagCTCGAGgcggactggtgc
(-)cCP-R agatACGCGTagcggtcagcgagttgg
(-d)CP:

(-)dCP-F ccagCTCGAGgctgcggactgg
(-)dCP-R agatACGCGTggtcagcgagttggtcatc

(-e)CP:
(-)eCP-F ccagCTCGAGaccgctgcggac
(-)eCP-R agatACGCGTcagcgagttggtcatcaca

(-f)CP:
(-)fCP-F ccagCTCGAGctgaccgctgcg
(-)fCP-R agatACGCGTcgagttggtcatcacaggacc

(-g)CP:
(-)gCP-F ccagCTCGAGtcgctgaccgct
(-)gCP-R agatACGCGTgttggtcatcacaggaccgt

(-h)CP:
(-)hCP-F ccagCTCGAGaactcgctgaccgc
(-)hCP-R agatACGCGTggtcatcacaggaccgtc

(-i)CP:
(-)iCP-F ccagCTCGAGaccaactcgctgaccg
(-)iCP-R agatACGCGTcatcacaggaccgtcagca

(-j)CP:

(-)jCP-F	ccagCTCGAGatgaccaactcgctgacc
(-)jCP-R	agatACGCGTcacaggaccgtcagcag

(-k)CP:

(-)kCP-F	ccagCTCGAGgtgatgaccaactcgctg
(-)kCP-R	agatACGCGTaggaccgtcagcaggga

(-l)CP:

(-)lCP-F	ccagCTCGAGcctgtgatgaccaactcg
(-)lCP-R	agatACGCGTaccgtcagcagggaaacc

(-m)CP:

(-)mCP-F	ccagCTCGAGggtcctgtgatgacca
(-)mCP-R	agatACGCGTgtcagcagggaaaccagtc

(-n)CP:

(-)nCP-F	ccagCTCGAGgacggtcctgtgatgac
(-)nCP-R	agatACGCGTagcagggaaaccagtccc

(20+fb)CP:

FW(20F+)bfCP_mNeon	ccagCTCGAGaccactggaaat
RV(20F+)bfCP_mNeon	agatACGCGTgtaactccacttaa

(10+fb):

FW(10F+)bfCP_mNeon	ccagCTCGAGaccatcatcagt
RV(10F+)bfCP_mNeon	agatACGCGTtttgtcgttggg

5.8 References

1. Zimmer, M. Green Fluorescent Protein (GFP): Applications , Structure , and Related Photophysical Behavior. (2002).
2. Youvan, D. *Green fluorescent pets*. *Science* **268,** (2006).
3. Baird, G. S., Zacharias, D. A. & Tsien, R. Y. Circular permutation and receptor insertion within green fluorescent proteins. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 11241–11246 (1999).
4. Kotlikoff, M. I. Genetically encoded Ca2+indicators: Using genetics and molecular design to understand complex physiology. *J. Physiol.* **578,** 55–67 (2007).
5. Zhang, J., Campbell, R. E., Ting, A. Y. & Tsien, R. Y. CREATING NEW FLUORESCENT PROBES FOR CELL BIOLOGY. **3,** 906–918 (2002).
6. Tantama, M., Martínez-François, J. R., Mongeon, R. & Yellen, G. Imaging energy status in live cells with a fluorescent biosensor of the intracellular ATP-to-ADP ratio. *Nat. Commun.* **4,** (2013).
7. Bilan, D. S. & Belousov, V. V. New tools for redox biology: From imaging to manipulation. *Free Radic. Biol. Med.* **109,** 167–188 (2017).
8. Zhao, Y. *et al.* Genetically Encoded Fluorescent Sensors for Intracellular NADH Detection. *Cell Metab.* **14,** 555–566 (2011).
9. Shaner, N. C. *et al.* A bright monomeric green fluorescent protein derived from Branchiostoma lanceolatum. *Nat Methods* **10,** 407–409 (2013).
10. Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499,** 295–300 (2013).

11. Guo, Q. *et al.* Multi-channel fiber photometry for population neuronal activity recording. *Biomed. Opt. Express* **6,** 3919 (2015).

12. Ouzounov, D. G. *et al.* In vivo three-photon imaging of activity of GCaMP6-labeled neurons deep in intact mouse brain. *Nat. Methods* **14,** 388–390 (2017).

13. J. Berg, Y. P. Hung, G. Y. A genetically encoded fluorescent reporter of ATP/ADP ratio. *Nat. Methods* **15,** 363–374 (2009).

14. Shui, B. *et al.* Circular permutation of red fluorescent proteins. *PLoS One* **6,** (2011).

15. Baird, G. S., Zacharias, D. A. & Tsien, R. Y. Circular permutation and receptor insertion within green fluorescent proteins. *Proc. Natl. Acad. Sci.* **96,** 11241–11246 (1999).

16. Carmody, W. R. An easily prepared wide range buffer series. *J. Chem. Educ.* **40,** A386 (2009).

17. Hoi, H. *et al.* An engineered monomeric zoanthus sp. yellow fluorescent protein. *Chem. Biol.* **20,** 1296–1304 (2013).

18. Steiert, F., Petrov, E. P., Schultz, P., Schwille, P. & Weidemann, T. Photophysical Behavior of mNeonGreen, an Evolutionarily Distant Green Fluorescent Protein. *Biophys. J.* **114,** 2419–2431 (2018).

19. Clavel, D. *et al.* Structural analysis of the bright monomeric yellow-green fluorescent protein mNeonGreen obtained by directed evolution. *Acta Crystallogr. Sect. D Struct. Biol.* **72,** 1298–1307 (2016).

20. Nagai, T. *et al.* A new design for a green calcium indicator with a smaller size and a reduced number of calcium-binding sites. *Protein Sci.* **6,** 1–15 (2016).